

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 09:30:02 ; Search time 6.00576 Seconds
(without alignments)
2914.068 Million cell updates/sec

Title: US-10-045-631B-88_COPY_2_340

Perfect score: 1744

Sequence: 1 AEGGAVCVVRPLNSREE.....QFASTAKYMKNTPYVNEVST 339

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA.*

- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pap.*
- 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pap.*
- 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pap.*
- 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pap.*
- 5: /cgn2_6/ptodata/2/iaa/PTUS_COMB.pap.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1744	100.0	2662	4	US-09-595-684B-31
2	1299.5	74.5	2954	4	US-09-150-867-1
3	668	38.3	473	4	US-09-592-054-6
4	668	38.3	522	4	US-09-592-054-4
5	658	37.7	1231	4	US-09-595-684B-23
6	652	37.4	1232	4	US-09-592-054-2
7	651	37.3	1234	4	US-09-592-054-8
8	615.5	35.3	411	2	US-08-713-815A-4
9	615.5	35.3	441	2	US-08-713-815A-3
10	615.5	35.3	975	4	US-09-914-259-19
11	615	35.3	409	4	US-09-592-191-6
12	615	35.3	409	4	US-09-723-262-6
13	615	35.3	409	4	US-09-723-219-6
14	615	35.3	1388	4	US-09-592-191-2
15	615	35.3	1388	4	US-09-723-262-2
16	615	35.3	1388	4	US-09-723-219-2
17	612.5	35.1	1031	4	US-09-914-259-24
18	605	34.7	1066	3	US-09-541-782-8
19	605	34.7	1066	4	US-09-723-820-8
20	605	34.7	1066	4	US-10-270-085-8
21	604	34.6	963	4	US-09-914-259-22
22	603	34.6	375	4	US-09-572-191-4
23	603	34.6	375	4	US-09-723-262-4
24	603	34.6	375	4	US-09-723-219-4
25	603	34.6	963	4	US-09-914-259-20
26	602.5	34.5	1032	4	US-09-914-259-26
27	600.5	34.4	1027	4	US-09-914-259-27

28	599	34.3	957	4	US-09-914-259-16	Sequence 16, Appl
29	598.5	34.3	967	4	US-09-914-259-21	Sequence 21, Appl
30	598	34.3	928	4	US-09-914-259-23	Sequence 23, Appl
31	595	34.1	956	4	US-09-914-259-17	Sequence 17, Appl
32	594.5	34.1	1279	4	US-09-724-517-2	Sequence 2, Appli
33	594.5	34.1	1279	4	US-09-641-807A-2	Sequence 2, Appli
34	594.5	34.1	1279	4	US-09-723-096-2	Sequence 2, Appli
35	591	33.9	815	4	US-09-914-259-18	Sequence 18, Appl
36	587.5	33.7	935	4	US-09-914-259-25	Sequence 25, Appl
37	581	33.3	1690	4	US-09-595-684B-39	Sequence 39, Appl
38	574.5	32.9	341	4	US-09-724-517-4	Sequence 4, Appli
39	574.5	32.9	341	4	US-09-641-807A-4	Sequence 4, Appli
40	574.5	32.9	341	4	US-09-723-096-4	Sequence 2, Appli
41	569.5	32.7	342	4	US-09-641-806-2	Sequence 2, Appli
42	569.5	32.7	342	4	US-09-723-129-2	Sequence 2, Appli
43	569.5	32.7	342	4	US-09-722-862-2	Sequence 2, Appli
44	569.5	32.7	1375	4	US-09-722-139-2	Sequence 2, Appli
45	569.5	32.7	1375	4	US-09-721-832-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-595-684B-31
; Sequence 31, Application US/09595684B
; Patent No. 6544766
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Ohashi, Cara
; APPLICANT: Sakowicz, Roman
; APPLICANT: Vaisberg, Eugeni
; APPLICANT: Wood, Kenneth
; APPLICANT: Yu, Ming
; TITLE OF INVENTION: Human kinesins and methods of producing
; FILE REFERENCE: cytop036
; CURRENT APPLICATION NUMBER: US/09/595,684B
; CURRENT FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: 09/295,612
; PRIOR FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 2662
; TYPE: PRT
; ORGANISM: Human
US-09-595-684B-31

Query Match 100.0%; Score 1744; DB 4; Length 2662;
Best Local Similarity 100.0%; Pred. No. 1.9e-176;
Matches 339; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	AEGGAVCVVRPLNSREESLGETAQVYKTDNNVYQVDSKSFNFDRVFGNETTKN	60
Db	2	AEGGAVCVVRPLNSREESLGETAQVYKTDNNVYQVDSKSFNFDRVFGNETTKN	61
Qy	61	VYEETIAPIIDSAIQGYNGTIFAYGQTAGCTYTMGSEDHGLGVIPRAIHDIQKIKKFP	120
Db	62	VYEETIAPIIDSAIQGYNGTIFAYGQTAGCTYTMGSEDHGLGVIPRAIHDIQKIKKFP	121
Qy	121	DREFLARVSMELYNETITDLCGTQKMKPLIREDVNRNRYVADLTTEEVYVYSEMAKWK	180
Db	122	DREFLARVSMELYNETITDLCGTQKMKPLIREDVNRNRYVADLTTEEVYVYSEMAKWK	181
Qy	181	ITKGEKSRHYGETKMQRSRSHITFRMILESREKGEPSNCEGSKVSHLNLVDLAGSER	240
Db	182	ITKGEKSRHYGETKMQRSRSHITFRMILESREKGEPSNCEGSKVSHLNLVDLAGSER	241
Qy	241	AAQTGAAGVRLKEGCNINRSLFILGVIKLSDGQVGGFINYRDSKLTRELQNSLGNPK	300
Db	242	AAQTGAAGVRLKEGCNINRSLFILGVIKLSDGQVGGFINYRDSKLTRELQNSLGNPK	301

```

QY 301 TRIICITIPVSGFDELTALQFASTAKYMKNTPYVNEVST 339
DB 302 TRIICITIPVSGFDELTALQFASTAKYMKNTPYVNEVST 340

RESULT 2
US-09-150-867-1
; Sequence 1, Application US/09150867
; Patent No. 6645748
; GENERAL INFORMATION:
; APPLICANT: Wood, Kenneth W.
; APPLICANT: Sakowicz, Roman
; APPLICANT: Goldstein, Lawrence S.B.
; APPLICANT: Cleveland, Don W.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Plus End-Directed Microtubule Motor Required for
; FILE REFERENCE: 18557C-000110US
; CURRENT APPLICATION NUMBER: US/09/150,867
; CURRENT FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: US 60/058,645
; EARLIER FILING DATE: 1997-09-11
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2954
; TYPE: PRT
; ORGANISM: Xenopus sp.
; FEATURE:
; OTHER INFORMATION: Xenopus centromere-associated protein-E (XCENP-E)
; OTHER INFORMATION: member of the kinesin superfamily of microtubule
; OTHER INFORMATION: motor proteins
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (1)..(472)
; OTHER INFORMATION: kinesin like motor domain
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (473)..(2752)
; OTHER INFORMATION: rod domain
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (2753)..(2954)
; OTHER INFORMATION: tail domain
; OTHER INFORMATION: tail domain
US-09-150-867-1

Query Match 74.5%; Score 1299.5; DB 4; Length 2954;
Best Local Similarity 73.5%; Pred. No. 7.3e-129; Indels 5; Gaps 2;
Matches 250; Conservative 42; Mismatches 43;

QY 1 AEGGAVAVCVVRPLNRSREESLGTAAQVYWKTDNNVIYQVDSKSNFDRVPHGNETTKN 60
DB 2 SEGDAVAVCVVRPLNRSREESLGTAAQVYWKTDNNVIYQVDSKSNFDRVPHGNETTKN 60
QY 61 VYEEIAPIIDSALQGYNGTIFAYGQTASGKTYTMGSEDLGVIPRAIHDFOKIKKFP 120
DB 60 IYQBIAPVIRALSALQGYNGTIFAYGQTSSGKTYTMGTPNSLGIPOAIOEVFKIIEIP 119
QY 121 DREPLLVSYMEIYNETITDLCCGQKMKPLIREDVNRNYYVADLTVEVYVTSMAALKW 180
DB 120 NREFLLVSYMEIYNETITDLCCGQKMKPLIREDVNRNYYVADLTVEVYVTSMAALKW 179
QY 181 ITKGEKSRHYGETKMNORSRSHITFRMILESREKGEPS---NCEGSKVSHLNLVDLAG 237
DB 180 IKKEKSRHYGETKMNDRSHRSHITFRMIVESRDRNDPTNSCNDGAVVSHLNLVDLAG 239
QY 238 SERAAQTGAAGVRLKEGCNINRSFILGQVTKKLSGQVGGFINYRDSKLTIRLQNSLGG 297
DB 240 SERASQTGAEGVRLKEGCNINRSFILGQVTKKLSGQVGGFINYRDSKLTIRLQNSLGG 299
QY 298 NPKTRIICITIPVSGFDELTALQFASTAKYMKNTPYVNEV 337
DB 300 NAKTVIICITIPVSGFDELTALQFASTAKYMKNTPYVNEV 339

```

```

RESULT 3
US-09-592-054-6
; Sequence 6, Application US/09592054
; Patent No. 6440684
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Finer, Jeffrey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Wood, Kenneth
; TITLE OF INVENTION: No. 6440684e1 motor proteins and methods for
; TITLE OF INVENTION: their use
; FILE REFERENCE: 1016
; CURRENT APPLICATION NUMBER: US/09/592,054
; CURRENT FILING DATE: 2000-07-20
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 473
; TYPE: PRT
; ORGANISM: Human
US-09-592-054-6

Query Match 38.3%; Score 668; DB 4; Length 473;
Best Local Similarity 44.7%; Pred. No. 1.3e-62;
Matches 152; Conservative 52; Mismatches 120; Indels 16; Gaps 5;

QY 6 VAVCVVRPLNRSREESLGTAAQVYWKTDNNVIYQVDSKSNFDRVPHGNETTKVYEE 64
DB 8 VRVALRCRPLVPKEISEGQCMCLSFVPGPQVVGVD--KSFTYDFVDFPSTQEVEVNT 65
QY 65 IAAPITDSALQGYNGTIFAYGQTASGKTYTMG-----SEDLGVIPRAIHDFOKIK 117
DB 66 AVAPLIKGVKGYNATVLAAYGQTSGKTYSMGAYTAEQNEFTVGVIPVLIQLLFXEID 125
QY 118 KPDRFELLRVSYMEIYNETITDLCCGQKMKPLIREDVNRNYYVADLTVEVYVTSMA 177
DB 126 KKSDFEFTLVSYLEIYNEIILLDCPSREKAGINIREDPKEGKIIVGLTEKTVLALDT 185
QY 178 LKWITKGEKSRHYGETKMNORSRSHITFRMILESREKGEPSCEGSKVSHLNLVDLAG 237
DB 186 VSCLEQGNNSRTVASTAMNSQSSRSHAFTISLEQRKKS-----KNSFSRSLHLVDLAG 241
QY 238 SERAAQTGAAGVRLKEGCNINRSFILGQVTKKLSGQVGGFINYRDSKLTIRLQNSLGG 297
DB 242 SERQKTKAEGDLKKEGININRGLCLGNVISALGDDKGGFVYRDSKLTIRLQNSLGG 301
QY 298 NPKTRIICITIPV--SFDELTALQFASTAKYMKNTPYVN 335
DB 302 NSHTLMIACVSPADSNLEETLNLRYADRARKINKKPIVN 341

RESULT 4
US-09-592-054-4
; Sequence 4, Application US/09592054
; Patent No. 6440684
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Finer, Jeffrey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Wood, Kenneth
; TITLE OF INVENTION: No. 6440684e1 motor proteins and methods for
; TITLE OF INVENTION: their use
; FILE REFERENCE: 1016
; CURRENT APPLICATION NUMBER: US/09/592,054
; CURRENT FILING DATE: 2000-07-20
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 522
; TYPE: PRT
; ORGANISM: Human

```


US-09-592-054-4

Query Match 38.3%; Score 668; DB 4; Length 522;
Best Local Similarity 44.7%; Pred. No. 1.5e-62;

Matches 152; Conservative 52; Mismatches 120; Indels 16; Gaps 5;

QY 6 VAVCVVRPLNRSRLG-ETAQVYWKTDNNVIYQVDGSKSFNDRVPHGNKTTKNVYEE 64

Db 30 VRVALRCRPLVKEISEGCMCLSFVPGEPQVVVGTD--KSFTYDFVDPSTQEVEFNT 87

QY 65 IAAPILDSATQYNGTIFAYGOTASGKTYTMG-----SEDHLGVIPRAIHDFQKIK 117

Db 88 AVAPLKGKGVFNATVLAIGTSGKTYSMGAYTAQENEPVGVIPRVQLLFKEID 147

QY 118 KFPDREFLLRVSYMEIYNETITDLCGTOKMKPLIREDVNRNVYVADLTEEVVYTSEMA 177

Db 148 KKSDFEFTLKVSYLEIYNEEILDLCPSEKQAQINREDPKGIKIVGLTEKTVLVALDT 207

QY 178 LKWITGKSRHYGRTKNNQSRSHITFRMILESEKGEPCNSCEGSKVSHNLVLDLAG 237

Db 208 VSCLEQGNNSRTVASTAMNSQSSRSHAITFTISLEQKKSD----KNSSFRSKLHLVLDLAG 263

QY 238 SERAAQTGAAGVRLKEGNCNINRSFLIGQVVKLSDGQVGGFYNRDSKLTIRLQNSLGG 297

Db 264 SERQKTKAEGDRLKEGININRGLLCLGNVISALGDDKGGFVYRDSKLTIRLQNSLGG 323

QY 298 NPKTRIICTITPV--SFDETLTALQFASHTAKYMKNTPVN 335

Db 324 NSHTLMIACVSPADSNLEETLNTLYADRAKIKNKPIVN 363

RESULT 5

US-09-595-684B-23

; Sequence 23, Application US/09595684B

; Patent No. 6544766

; GENERAL INFORMATION:

; APPLICANT: Beraud, Christophe

; APPLICANT: Ohashi, Cara

; APPLICANT: Sakowicz, Roman

; APPLICANT: Vaisberg, Eugeni

; APPLICANT: Wood, Kenneth

; APPLICANT: Yu, Ming

; TITLE OF INVENTION: Human kinesins and methods of producing

; FILE OF INVENTION: and purifying human kinesins

; FILE REFERENCE: cytop036

; CURRENT APPLICATION NUMBER: US/09/595,684B

; CURRENT FILING DATE: 2002-06-24

; PRIOR APPLICATION NUMBER: 09/295,612

; PRIOR FILING DATE: 2000-04-20

; NUMBER OF SEQ ID NOS: 105

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 23

; LENGTH: 1231

; TYPE: PRT

; ORGANISM: Human

US-09-595-684B-23

Query Match 37.7%; Score 658; DB 4; Length 1231;

Best Local Similarity 44.7%; Pred. No. 7e-61;

Matches 152; Conservative 52; Mismatches 120; Indels 16; Gaps 6;

QY 6 VAVCVVRPLNRSRLG-ETAQVYWKTDNNVIYQVDGSKSFNDRVPHGNKTTKNVYEE 64

Db 10 VRVALRCRPLVKEISEGCMCLSFVPGEPQVVVGTD--KSFTYDFVDPSTQEVEFNT 67

QY 65 IAAPILDSATQYNGTIFAYGOTASGKTYTMG-----SEDHLGVIPRAIHDFQKIK 117

Db 68 AVAPLKGKGVFNATVLAIGTSGKTYSMGAYTAQENEPVGVIPRVQLLFKEID 127

QY 118 KFPDREFLLRVSYMEIYNETITDLCGTOKMKPLIREDVNRNVYVADLTEEVVYTSEMA 177

Db 128 KKSDFEFTLKVSYLEIYNEEILDLCPSEKQAQINREDPKGIKIVGLTEKTVLVALDT 187

QY 178 LKWITGKSRHYGRTKNNQSRSHITFRMILESEKGEPCNSCEGSKVSHNLVLDLAG 237

Db 188 VSCLEQGNNSRTVASTAMNSQSSRSHAITFTISLEQKKSDKNSSFR-SKLHLVLDLAG 243

QY 238 SERAAQTGAAGVRLKEGNCNINRSFLIGQVVKLSDGQVGGFYNRDSKLTIRLQNSLGG 297

Db 244 SERQKTKAEGDRLKEGININRGLLCLGNVISALGDDKGGFAPYRDSKLTIRLQNSLGG 303

QY 298 NPKTRIICTITPV--SFDETLTALQFASHTAKYMKNTPVN 335

Db 304 NSHTLMIACVSPADSNLEETLNTLYADRAKIKNKPIVN 343

RESULT 6

US-09-592-054-2

; Sequence 2, Application US/09592054

; Patent No. 6440684

; GENERAL INFORMATION:

; APPLICANT: Beraud, Christophe

; APPLICANT: Finer, Jeffrey

; APPLICANT: Sakowicz, Roman

; APPLICANT: Wood, Kenneth

; TITLE OF INVENTION: No. 6440684el motor proteins and methods for

; FILE OF INVENTION: their use

; FILE REFERENCE: 1016

; CURRENT APPLICATION NUMBER: US/09/592,054

; CURRENT FILING DATE: 2000-07-20

; NUMBER OF SEQ ID NOS: 8

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 2

; LENGTH: 1232

; TYPE: PRT

; ORGANISM: Human

US-09-592-054-2

Query Match 37.4%; Score 652; DB 4; Length 1232;

Best Local Similarity 44.4%; Pred. No. 3.1e-60;

Matches 151; Conservative 52; Mismatches 121; Indels 16; Gaps 6;

QY 6 VAVCVVRPLNRSRLG-ETAQVYWKTDNNVIYQVDGSKSFNDRVPHGNKTTKNVYEE 64

Db 10 VRVALRCRPLVKEISEGCMCLSFVPGEPQVVVGTD--KSFTYDFVDPSTQEVEFNT 67

QY 65 IAAPILDSATQYNGTIFAYGOTASGKTYTMG-----SEDHLGVIPRAIHDFQKIK 117

Db 68 AVAPLKGKGVFNATVLAIGTSGKTYSMGAYTAQENEPVGVIPRVQLLFKEID 127

QY 118 KFPDREFLLRVSYMEIYNETITDLCGTOKMKPLIREDVNRNVYVADLTEEVVYTSEMA 177

Db 128 KKSDFEFTLKVSYLEIYNEEILDLCPSEKQAQINREDPKGIKIVGLTEKTVLVALDT 187

QY 178 LKWITGKSRHYGRTKNNQSRSHITFRMILESEKGEPCNSCEGSKVSHNLVLDLAG 237

Db 188 VSCLEQGNNSRTVASTAMNSQSSRSHAITFTISLEQKKSDKNSSFR-SKLHLVLDLAG 243

QY 238 SERAAQTGAAGVRLKEGNCNINRSFLIGQVVKLSDGQVGGFYNRDSKLTIRLQNSLGG 297

Db 244 SERQKTKAEGDRLKEGININRGLLCLGNVISALGDDKGGFAPYRDSKLTIRLQNSLGG 303

QY 298 NPKTRIICTITPV--SFDETLTALQFASHTAKYMKNTPVN 335

Db 304 NSHTLMIACVSPADSNLEETLNTLYADRAKIKNKPIVN 343

RESULT 7

US-09-592-054-8

; Sequence 8, Application US/09592054

; Patent No. 6440684

; GENERAL INFORMATION:

; APPLICANT: Beraud, Christophe

; APPLICANT: Finer, Jeffrey

; APPLICANT: Sakowicz, Roman

; APPLICANT: Wood, Kenneth

;; TITLE OF INVENTION: No. 6440684e1 motor proteins and methods for

;; FILE REFERENCE: 1016

;; CURRENT APPLICATION NUMBER: US/09/592,054

;; CURRENT FILING DATE: 2000-07-20

;; NUMBER OF SEQ ID NOS: 8

;; SOFTWARE: FastSeq for Windows Version 4.0

;; SEQ ID NO 8

;; LENGTH: 1234

;; TYPE: PRT

;; ORGANISM: Human

US-09-592-054-8

Query Match 37.3%; Score 651; DB 4; Length 1234;
Best Local Similarity 43.8%; Pred. No. 3.9e-60;
Matches 149; Conservative 56; Mismatches 119; Indels 16; Gaps 6;
QY 6 VAYCVVRPLNSREESIG-ETAQVYWKTDNNVIYQVDSKSFNDRVFGHNETKNVYEE 64
DB 10 VRVALCRPLVPKEISEGCMCLSFVPGETQVVVGTGDKSFTYDFVDFPCTEQEERFVK 67
QY 65 IAAPIIDSALQYNGTIFAYGQTASGKTYTMG-----SEDLHGVIPRAIHDFQKIK 117
DB 68 AVAPLIXGIPKYNATVLAAYGOTSGKTSYMGAYTAEQENETVGLIPRVIQLLFKKID 127
QY 118 KPDPRELLRVSYMEIYNETITDLGCTQKMKPLIREDVNRNIVYADLITEEVVYTSMA 177
DB 128 QKSDFFETLVSVLEIYNEELDLCPREKAQINREDPKGKIVGLTEKTVLVALDT 187
QY 178 LKWTITGKSKRHYGETKMNORSRSHTIFPMILESRKGPSPNCEGSKVSHNLVDLAG 237
DB 188 VSCLEQNNSETVASTAMNQSRSRSHAFITII--SLEQKKSKDNSSFR-SKLHLVDLAG 243
QY 238 SERAAQTGAAGVRLKEGCNINRSFILGQVTKLSDGVGFGFNYRDSKLTIRLQNSLGG 297
DB 244 SERQKTKARGDLKEGININRGLCLGNVISALGDDKGSFVYRDSKLTIRLQNSLGG 303
QY 298 NPKTRIICITPV--SPDELTALQFASHTAKYMKNTPVYN 335
DB 304 NSHTLMACVSPADSNLEETLSTLRADRAKINKKIVN 343

RESULT 8

US-08-713-815A-4

;; Sequence 4, Application US/08/713815A

;; Patent No. 5830659

;; GENERAL INFORMATION:

;; APPLICANT: Russell J. Stewart

;; TITLE OF INVENTION: ACTIVE MICROTUBULE-BASED

;; TITLE OF INVENTION: SEPARATIONS BY KINESINS

;; NUMBER OF SEQUENCES: 6

;; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: Thorpe, No. 5830659th & Western, L.L.P.

;; STREET: 9035 South 700 East, Suite 200

;; CITY: Sandy

;; STATE: Utah

;; COUNTRY: USA

;; ZIP: 84070

;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage

;; COMPUTER: AST Ascentia 900N

;; OPERATING SYSTEM: DOS 6.22

;; SOFTWARE: Word Perfect 6.0

;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/08/713,815A

;; FILING DATE: 13-SEP-1996

;; CLASSIFICATION: 435

;; ATTORNEY/AGENT INFORMATION:

;; NAME: Alan J. Howarth

;; REGISTRATION NUMBER: 36,553

;; REFERENCE/DOCKET NUMBER: T3214/U-2202

;; TELECOMMUNICATION INFORMATION:

;; TELEFAX: (801)566-0750

;; INFORMATION FOR SEQ ID NO: 4:

;; SEQUENCE CHARACTERISTICS:

;; LENGTH: 411 amino acid residues

;; TYPE: amino acid

;; TOPOLOGY: linear

US-08-713-815A-4

Query Match 35.3%; Score 615.5; DB 2; Length 411;

Best Local Similarity 45.3%; Pred. No. 4.2e-57;

Matches 158; Conservative 43; Mismatches 125; Indels 23; Gaps 12;

QY 1 AREGAVAVCVVRPLNSREESIGETAQVYWKTDNNV---IYQVDSKSFNDRVFGHNET 57
DB 9 AEDSIKVCV-RRPLNDSEKAGSKFVV--KFPNNVEENCISLAG-KVYLFDKVFKPNAS 64
QY 58 TKNVVEEIAAPITDSALQYNGTIFAYGQTASGKTYTM---MGSEDLHGVIPRAIHDFQ 114
DB 65 QEKVYNEAKSIVTDVLAAGYNGTIFAYGQTSSGKTHMEGVIGDSVKQGIIPRIVNDIFN 124
QY 115 KIKKFP-DREFLRVSYMEIYNETITDLGCTQKMKPLIREDVNRNIVYADLITEEVVYT 173
DB 125 HIYAMEVNLFEHIKVSYYEIMDKIRDLL-DVSKVN-LSVHEDKNRPVYKGAFTERVSS 182
QY 174 SEMALKWITKGEKSRHYGETKMNORSRSHTIFPMILESRKGPSPNCEGSKVSHNL 232
DB 183 PEDVFEIIEGKGNRHIAVNNHESRSRSHSVFLINVQ-----ENLENQKLSGKLYL 236
QY 233 VDLAGSERAQTGAAGVRLKEGCNINRSFILGQVTKLSDGVGFGFNYRDSKLTIRLQ 292
DB 237 VDLAGSEKVSCTGAEVTDLDEAKNINKLSALGNVISALADGN-KTHIPYRDSKLTIRLQ 295
QY 293 NSLGGNPKTRIICITPVSPD--ETLTALQFASHTAKYMKNTPVYNEVST 339
DB 296 ESLGNGARTTIVCCSPASFNESETKSLDGFRRAKTVKRVNVCVNEELT 344

RESULT 9

US-08-713-815A-3

;; Sequence 3, Application US/08/713815A

;; Patent No. 5830659

;; GENERAL INFORMATION:

;; APPLICANT: Russell J. Stewart

;; TITLE OF INVENTION: ACTIVE MICROTUBULE-BASED

;; TITLE OF INVENTION: SEPARATIONS BY KINESINS

;; NUMBER OF SEQUENCES: 6

;; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: Thorpe, No. 5830659th & Western, L.L.P.

;; STREET: 9035 South 700 East, Suite 200

;; CITY: Sandy

;; STATE: Utah

;; COUNTRY: USA

;; ZIP: 84070

;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage

;; COMPUTER: AST Ascentia 900N

;; OPERATING SYSTEM: DOS 6.22

;; SOFTWARE: Word Perfect 6.0

;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/08/713,815A

;; FILING DATE: 13-SEP-1996

;; CLASSIFICATION: 435

;; ATTORNEY/AGENT INFORMATION:

;; NAME: Alan J. Howarth

;; REGISTRATION NUMBER: 36,553

;; REFERENCE/DOCKET NUMBER: T3214/U-2202

;; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: (801)566-6633

;; TELEFAX: (801)566-0750

;; INFORMATION FOR SEQ ID NO: 3:

;; SEQUENCE CHARACTERISTICS:

;; LENGTH: 441 amino acid residues

;; TYPE: amino acid


```

; CURRENT APPLICATION NUMBER: US/09/723,262
; CURRENT FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: US 09/572,191
; PRIOR FILING DATE: 2000-05-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Human
; US-09-723-262-6

Query Match          35.3%; Score 615; DB 4; Length 409;
Best Local Similarity 43.3%; Pred. No. 4.8e-57;
Matches 158; Conservative 48; Mismatches 113; Indels 46; Gaps 10;

QY 2 EGAVALVVRPLNSREESL-GETAQVYWKTDNNVIYQVDGS-----KSFNFD 49
Db 23 EGAVALVVRIRPPAERSGSADGE-----QNLCLSVLSSTSLRLHSNPEPKTFTFD 73

QY 50 RVPHGNETTKNVEEIAAPIIDSAIQYNGTIFAYGQTASGKTYTMMG-----SEDHL 102
Db 74 HVADVDVTQESVFATVAKSIVSCMSGYNGTIFAYGQTGSGKTFMMGPSESNFHNLR 133

QY 103 GVIPRAIHDF-----QKIKFPDPREFLLRVSYMEIYNETITDLCGTQKMKPLIREDV 157
Db 134 GVIPRSFEYLFSLIDREKEKAGAKSFLCKSCFIEIYNEQIYDLL--DSASAGLYLREHI 191

QY 158 NRVVYVADLTVEVYVYSEMALKWITKGEKSRHVGKTMNQRSSRSHTIFFMILESREKGE 217
Db 192 KKGVFVVGAVEQVVTSAEAYQVLGGWRNRVASTSMNRESRSSHAVFTTIESMEK-- 249

QY 218 PSNCEGSKVSHLNLVDLAGSRAAQTGAAGVRLKEGCNINRSLFLILGOVVKLSGQVVG 277
Db 250 -SNEIVNIRTSLNLVDLAGSERQKDTAEGMRLKEAGNINRSLSCLGQVITALVD--VG 306

QY 278 G-----FINVRDSKLTILQNSLGGNPKTRIICTITPVS--FDETLTALQFASAKYMKNT 331
Db 307 NGKQHVCYRDSKLTFLRLDSLGGNAKTAIIANVHPGSRFCFGETLSTLNPAQAKLIKKN 366

QY 332 PYNE 336
Db 367 AVNE 371

RESULT 14
US-09-572-191-2
; Sequence 2, Application US/09572191
; Patent No. 6355466
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Sakowicz, Roman
; APPLICANT: Wood, Kenneth
; TITLE OF INVENTION: No. 6355466el motor proteins and methods for
; FILE OF INVENTION: their use
; FILE REFERENCE: 1017
; CURRENT APPLICATION NUMBER: US/09/572,191
; CURRENT FILING DATE: 2000-05-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1388
; TYPE: PRT
; ORGANISM: Human
; US-09-572-191-2

Query Match          35.3%; Score 615; DB 4; Length 1388;
Best Local Similarity 43.3%; Pred. No. 3.4e-56;
Matches 158; Conservative 48; Mismatches 113; Indels 46; Gaps 10;

QY 2 EGAVALVVRVPLNSREESL-GETAQVYWKTDNNVIYQVDGS-----KSFNFD 49
Db 23 EGAVALVVRIRPPAERSGSADGE-----QNLCLSVLSSTSLRLHSNPEPKTFTFD 73

QY 50 RVPHGNETTKNVEEIAAPIIDSAIQYNGTIFAYGQTASGKTYTMMG-----SEDHL 102
Db 74 HVADVDVTQESVFATVAKSIVSCMSGYNGTIFAYGQTGSGKTFMMGPSESNFHNLR 133

QY 103 GVIPRAIHDF-----QKIKFPDPREFLLRVSYMEIYNETITDLCGTQKMKPLIREDV 157
Db 134 GVIPRSFEYLFSLIDREKEKAGAKSFLCKSCFIEIYNEQIYDLL--DSASAGLYLREHI 191

QY 158 NRVVYVADLTVEVYVYSEMALKWITKGEKSRHVGKTMNQRSSRSHTIFFMILESREKGE 217
Db 192 KKGVFVVGAVEQVVTSAEAYQVLGGWRNRVASTSMNRESRSSHAVFTTIESMEK-- 249

QY 218 PSNCEGSKVSHLNLVDLAGSRAAQTGAAGVRLKEGCNINRSLFLILGOVVKLSGQVVG 277
Db 250 -SNEIVNIRTSLNLVDLAGSERQKDTAEGMRLKEAGNINRSLSCLGQVITALVD--VG 306

; CURRENT APPLICATION NUMBER: US/09/723,262
; CURRENT FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: US 09/572,191
; PRIOR FILING DATE: 2000-05-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Human
; US-09-723-262-6

Query Match          35.3%; Score 615; DB 4; Length 409;
Best Local Similarity 43.3%; Pred. No. 4.8e-57;
Matches 158; Conservative 48; Mismatches 113; Indels 46; Gaps 10;

QY 2 EGAVALVVRPLNSREESL-GETAQVYWKTDNNVIYQVDGS-----KSFNFD 49
Db 23 EGAVALVVRIRPPAERSGSADGE-----QNLCLSVLSSTSLRLHSNPEPKTFTFD 73

QY 50 RVPHGNETTKNVEEIAAPIIDSAIQYNGTIFAYGQTASGKTYTMMG-----SEDHL 102
Db 74 HVADVDVTQESVFATVAKSIVSCMSGYNGTIFAYGQTGSGKTFMMGPSESNFHNLR 133

QY 103 GVIPRAIHDF-----QKIKFPDPREFLLRVSYMEIYNETITDLCGTQKMKPLIREDV 157
Db 134 GVIPRSFEYLFSLIDREKEKAGAKSFLCKSCFIEIYNEQIYDLL--DSASAGLYLREHI 191

QY 158 NRVVYVADLTVEVYVYSEMALKWITKGEKSRHVGKTMNQRSSRSHTIFFMILESREKGE 217
Db 192 KKGVFVVGAVEQVVTSAEAYQVLGGWRNRVASTSMNRESRSSHAVFTTIESMEK-- 249

QY 218 PSNCEGSKVSHLNLVDLAGSRAAQTGAAGVRLKEGCNINRSLFLILGOVVKLSGQVVG 277
Db 250 -SNEIVNIRTSLNLVDLAGSERQKDTAEGMRLKEAGNINRSLSCLGQVITALVD--VG 306

; CURRENT APPLICATION NUMBER: US/09723219
; CURRENT FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: US 09/572,191
; PRIOR FILING DATE: 2000-05-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Human
; US-09-723-219-6

Query Match          35.3%; Score 615; DB 4; Length 409;
Best Local Similarity 43.3%; Pred. No. 4.8e-57;
Matches 158; Conservative 48; Mismatches 113; Indels 46; Gaps 10;

QY 2 EGAVALVVRVPLNSREESL-GETAQVYWKTDNNVIYQVDGS-----KSFNFD 49
Db 23 EGAVALVVRIRPPAERSGSADGE-----QNLCLSVLSSTSLRLHSNPEPKTFTFD 73

QY 50 RVPHGNETTKNVEEIAAPIIDSAIQYNGTIFAYGQTASGKTYTMMG-----SEDHL 102
Db 74 HVADVDVTQESVFATVAKSIVSCMSGYNGTIFAYGQTGSGKTFMMGPSESNFHNLR 133

QY 103 GVIPRAIHDF-----QKIKFPDPREFLLRVSYMEIYNETITDLCGTQKMKPLIREDV 157
Db 134 GVIPRSFEYLFSLIDREKEKAGAKSFLCKSCFIEIYNEQIYDLL--DSASAGLYLREHI 191

QY 158 NRVVYVADLTVEVYVYSEMALKWITKGEKSRHVGKTMNQRSSRSHTIFFMILESREKGE 217
Db 192 KKGVFVVGAVEQVVTSAEAYQVLGGWRNRVASTSMNRESRSSHAVFTTIESMEK-- 249

QY 218 PSNCEGSKVSHLNLVDLAGSRAAQTGAAGVRLKEGCNINRSLFLILGOVVKLSGQVVG 277
Db 250 -SNEIVNIRTSLNLVDLAGSERQKDTAEGMRLKEAGNINRSLSCLGQVITALVD--VG 306

; CURRENT APPLICATION NUMBER: US/09723219
; CURRENT FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: US 09/572,191
; PRIOR FILING DATE: 2000-05-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Human
; US-09-723-219-6

Query Match          35.3%; Score 615; DB 4; Length 409;
Best Local Similarity 43.3%; Pred. No. 4.8e-57;
Matches 158; Conservative 48; Mismatches 113; Indels 46; Gaps 10;

QY 2 EGAVALVVRVPLNSREESL-GETAQVYWKTDNNVIYQVDGS-----KSFNFD 49
Db 23 EGAVALVVRIRPPAERSGSADGE-----QNLCLSVLSSTSLRLHSNPEPKTFTFD 73

QY 50 RVPHGNETTKNVEEIAAPIIDSAIQYNGTIFAYGQTASGKTYTMMG-----SEDHL 102
Db 74 HVADVDVTQESVFATVAKSIVSCMSGYNGTIFAYGQTGSGKTFMMGPSESNFHNLR 133

QY 103 GVIPRAIHDF-----QKIKFPDPREFLLRVSYMEIYNETITDLCGTQKMKPLIREDV 157
Db 134 GVIPRSFEYLFSLIDREKEKAGAKSFLCKSCFIEIYNEQIYDLL--DSASAGLYLREHI 191

QY 158 NRVVYVADLTVEVYVYSEMALKWITKGEKSRHVGKTMNQRSSRSHTIFFMILESREKGE 217
Db 192 KKGVFVVGAVEQVVTSAEAYQVLGGWRNRVASTSMNRESRSSHAVFTTIESMEK-- 249

QY 218 PSNCEGSKVSHLNLVDLAGSRAAQTGAAGVRLKEGCNINRSLFLILGOVVKLSGQVVG 277
Db 250 -SNEIVNIRTSLNLVDLAGSERQKDTAEGMRLKEAGNINRSLSCLGQVITALVD--VG 306
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 09:39:47 ; Search time 16.9319 Seconds

(without alignments)
6280.361 Million cell updates/sec

Title: US-10-045-631B-88_COPY_2_340

Perfect score: 1744

Sequence: 1 AEEGAVAVCVRVRLNSREE.....QPASTAKYMKNTPYNEVST 339

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US05_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/ECTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09D_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09E_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10D_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	829.5	47.6	1382	16	US-10-437-963-176714
2	793.5	45.5	694	12	US-10-425-114-59725
3	668	38.3	1232	14	US-10-116-712-670
4	668	38.3	1232	15	US-10-408-765A-2153
5	668	38.3	1235	16	US-10-334-143-8
6	666	38.2	1006	16	US-10-437-963-147933
7	664	38.1	956	16	US-10-437-963-182113
8	663.5	38.0	955	12	US-10-425-114-62672
9	663	38.0	1237	15	US-10-334-143-33
10	658	37.7	1232	14	US-10-116-712-664
11	658	37.7	1232	14	US-10-116-712-669
12	649.5	37.2	1029	16	US-10-311-642-2
13	647.5	37.1	420	16	US-10-437-963-114373
14	630	36.1	672	16	US-10-408-765A-1664
15	622.5	35.7	965	16	US-10-437-963-169150

16	620.5	35.6	329	16	US-10-311-642-4
17	619.5	35.5	966	12	US-10-425-114-62697
18	615.5	35.3	975	14	US-10-080-608A-19
19	615.5	35.3	975	15	US-10-370-685-108
20	615	35.3	409	12	US-10-332-089-6
21	615	35.3	1388	12	US-10-332-089-2
22	615	35.3	1388	14	US-10-146-473-82
23	615	35.3	1388	15	US-10-173-999-32
24	615	35.3	1388	16	US-10-188-832-164
25	612.5	35.1	1031	14	US-10-080-608A-24
26	612.5	35.1	1031	15	US-10-370-685-113
27	609	34.9	1045	16	US-10-437-963-150217
28	605.5	34.7	1189	12	US-10-424-599-258016
29	604	34.6	963	14	US-10-080-608A-22
30	604	34.6	963	15	US-10-370-685-111
31	603.5	34.6	757	12	US-10-220-120-366
32	603.5	34.6	757	16	US-10-363-829-316
33	603	34.6	375	12	US-10-332-089-4
34	603	34.6	963	14	US-10-080-608A-20
35	603	34.6	963	15	US-10-370-685-109
36	602.5	34.5	1011	12	US-10-336-472-24
37	602.5	34.5	1011	12	US-10-236-417-58
38	602.5	34.5	1032	14	US-10-080-608A-26
39	602.5	34.5	1123	12	US-10-370-685-115
40	602.5	34.5	1123	12	US-10-424-599-228690
41	600.5	34.4	1027	14	US-10-080-608A-27
42	600.5	34.4	1027	15	US-10-370-685-116
43	599	34.3	957	14	US-10-080-608A-16
44	599	34.3	957	15	US-10-370-685-105
45	598.5	34.3	517	15	US-10-369-493-12881

ALIGNMENTS

RESULT 1

US-10-437-963-176714
; Sequence 176714, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 176714
; LENGTH: 1382
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MBT4530_74437C.1.pap
US-10-437-963-176714

Query Match 47.6%; Score 829.5; DB 16; Length 1382;

Best Local Similarity 54.2%; Pred. No. 1.5e-75; Mismatches 95; Indels 17; Gaps 8;
Matches 186; Conservative 45;

QY 6 VAVCVRVRLNSRESLGETAQVYKTDNNVT-YQVDGSKSFNFRVHGNFTTKNVVEE 64

DB 4 IHVAVRAPLTA--EDAGSSP---WRVSGNALALSTQPSIRFEFRIEGEECTADVGA 58

QY 65 TAAPIIDSAIOGYNNTTIFAYGQTASGKTYTMWGSBDHIGVIPRAHDIFQTKKFPDREF 124

DB 59 RTKHIVDSAVRGNGTVFAYGQTNISGKTYTVRGSNGEPIIPLAVHDLFRTIEHLDREF 118


```

QY 125 LLRVSYMEIYNETITDLLCGTQKMKPLIIREDVNRNVVADLTBEVVYTSWALKWITKG 184
Db 119 LLRMSYMEIYNEEINDLL--VPEHRKQLIHESIERYIVAGLRBEIVTQPEQVLEFMSFG 176
QY 185 EKSRHGETKMNQSSPSHTIFRMLSSREKGEPS---NCEGSKVKGSHLNLVDLAGSER 240
Db 177 ESRHHIGETNNVYSSRSHITFRWVIESREKVDESEAGESCD-AVRVSVLNLVDLAGSER 235
QY 241 AAQTGAAGVRIKEGCNINRSFLILQGVIKKLSDG--QVGGFINYRDSKLTILQNSLGN 298
Db 236 AAKTGAEGVRLKEGSHINKSLMTLGTVIKKLSEGTEGGQGHVPYRDSKLTILQPALGN 295
QY 299 PKPRIICIT--PYSFDETLTALQFASHTAKYMKNTPYNEVST 339
Db 296 ANTAIIICITLAQVHADETKSSLOFASRALRVITNCACVNEILT 338

RESULT 2
US-10-425-114-59725
; Sequence 59725, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 59725
; LENGTH: 694
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3912-011-C4_FLI.pep
; US-10-425-114-59725

```

```

US-10-116-712-670
; Sequence 670, Application US/10116712
; Publication No. US20030194764A1
; GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Switzer, Ann
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.568
; CURRENT APPLICATION NUMBER: US/10/116,712
; CURRENT FILING DATE: 2002-04-07
; NUMBER OF SEQ ID NOS: 670
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 670
; LENGTH: 1232
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-116-712-670

Query Match      38.3%; Score 668; DB 14; Length 1232;
Best Local Similarity 44.7%; Pred. No. 5.6e-59;
Matches 152; Conservative 52; Mismatches 120; Indels 16; Gaps 5

Qy      6 VAVCVVRPLRRSRESLG-ETAQVYVKTDDNNVLYQVDSKSFNDRFVPHGNETTKNVYEE 64
Db      10 VRVALRCRPLVPKSIEISGCQMCLEFVGPGEQVVGTD--KSFTYDFVDPSTQEVEVT 67

Qy      65 IAAPIIDSAIQYNGTIFAYGQTASGKTYTMMG-----SEDLGVIPRAIHDFQIK 117
Db      68 AVAPLIKGVFKYGNATVLAGYQTSKTKYSMGGAYTAEQENEPTVGVIPRVIQLLFKEID 127

Qy      118 KFPDREFLLRYSYMEIYNETITDLLCGTKMKPLIREDNRNVVYADLTSEWVYTSEMA 177
Db      128 KSDPEFTLVKYSYLEIYNEELDLCPSREKAQINIREDPKEGIVGTEKTVLVALDT 187

Qy      178 LKWTITKGEKSRHYGETKNORSSRSHTIFRMILSREKGEPSNCEGSKVYSHNLIVDLAG 237
Db      188 VSCLEQGNRSRTVASTAMNSQSSRSHAFTTISLEQRKESD---KNSSFRSKLHLVDLAG 243

Qy      238 SERAAOTCAAGVRUKEGNCINRSFILGQVTKKISDGOVGGFINRYRDSKLTILQNSLGG 297
Db      244 SERQKTKAEGDRUKEGINRGLLCUNVISALGDDKKGFVYRDSKLTILLQDSLGG 303

Qy      298 NPKTRIICTTFV--SEDETLLAQFASTAKYMKNTPYVN 335
Db      304 NSHTIMTACVSPADSNLEETLNTLRYADRAKIKNKPVN 343

```

RESULT 3

Query Match	38.3%	Score 668;	DB 16;	Length 1232;
Best Local Similarity	44.7%;	Pred. No. 5.6e-59;		
Matches	152;	Conservative 52;	Mismatches 120;	Indels 16; Gaps 5;
QY	6	VAVCVVRPLNSRRESLG-ETAQVYWKTDNNVIYQVGSKFNRDVRVFGHNETTKNYYEE	64	
DB	10	VRVALRCRPLVFKISBGCMLGFVPGEPOVVGTD--KSFTYDFVDPSPTEQEEVENT	67	
QY	65	IAAPLIIDSAIGYNGTTFAYQQTASGKTTYTWMG-----SEDHLGVIPRAIHIDIFQIK	117	
DB	68	AVAPLLIKGVPGKYNATLAYQOTSGSKTYSMGGAYTAEOENPTGVIPRVLIQLLFKEID	127	
QY	118	KFPDRFLLRVSYMEIYNETITDLLCGTQKMKPLIIRDVNRNVVADLTBEVVYTSEMA	177	
DB	128	KKSDFEFTLVKSYLEIYNEETILDLLCPSRKAQINIREDPKEGIKVGLTEKTVLVALDIT	187	
QY	178	LKWITKEKSRHYGETKMNQSSRSHTIFRMILSRKEGFSNCEGSKVYSHLNAVLDLAG	237	
DB	188	VSCLEQGNNSRTVASTAMNQSQRSHAFTTSLBQRKSD---KNSSFRSKLHLVDLAG	243	
QY	238	SERAAOTGAAGVRUKEGCNINRSIFILGQVTKKLSDGQVGGFYINRYDRSKLTRIQLNSLGG	297	
DB	244	SERQKTKAEGDRUKIGINRGLLCGNVLSALGDGDKGFGVPYRDSKLTFLQLQDSLGG	303	
QY	298	NPKTRIICTIPV--SFEDETITALQFASTAKYMKNTPIYVN	335	
DB	304	NSHTLMITACVSPADNLEETLNTLRYADRASKIKNKPIVN	343	

```

RESULT 5
US-10-334-143-8
; Sequence 8, Application US/10334143
; Publication No. US2004009549A1
; GENERAL INFORMATION:
; APPLICANT: GRIGORIEV, IGOR VYACHESLAVOVICH
; APPLICANT: SUDARSANAM, SUCHA
; TITLE OF INVENTION: METHOD FOR DETECTING REMOTE HOMOLOGUES AND NOVEL
; TITLE OF INVENTION: KINASES IDENTIFIED WITH THE METHOD
; FILE REFERENCE: 038602/1543
; CURRENT APPLICATION NUMBER: US/10/334,143
; CURRENT FILING DATE: 2002-12-31
; PRIOR APPLICATION NUMBER: 60/343,169
; PRIOR FILING DATE: 2001-12-31
; NUMBER OF SEQ ID NOS: 207
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 1235
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-334-143-8

```

```

Db      247  SERQKTKAEGDRUKBGININRGLLCLGNVISALGDDKKGFPYPRDSKLTLLQLDSLGG 306

Qy      298  NPKTRIICITPPV--SFDETLTALQFASTAKYMKNTPIYN 335
      | | | | | : | | | | | : | | | | |
Db      307  NSHTLMTACVSPADSNLEETLNTLRYADRARAKIKKPIVN 346

RESULT 6
US-10-437-963-147933
; Sequence 147933, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 147933
; LENGTH: 1006
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1006)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_48414C.1.pep
US-10-437-963-147933

```

RESULT 7
US-10-437-963-182113
; Sequence 182113, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.

```
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 182113
; LENGTH: 956
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(956)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_79331C.1.pep
; US-10-437-963-182113
```

```
Query Match 38.1%; Score 664; DB 16; Length 956;
Best Local Similarity 46.2%; Pred. No. 9.8e-59;
Matches 160; Conservative 52; Mismatches 112; Indels 22; Gaps 8;

Qy 2 BEGAVAVCVVRPLNSREESLGETAQVYWKDNN--VYQVD-----GSKSPNDRVPHG 54
Db 31 KEEKIFVTVRPLSKKELAKD--QVWECDDNQTLTKYKPPQDRAAPTSTYFDKVFGP 88
Qy 55 NETTKNVEIEAIIPIIDSAIOGYNGTIFAYGQTASGKTYTMMGSEDLHGVIPRAIHDFQ 114
Db 89 ASQTEVVEEGAKDVAMSALTGINATIFAYGQTSSTGKTFTMR-----GVTESAVNDIYR 142
Qy 115 KIRKPPDRFLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNRYVADLTVEEVYTS 174
Db 143 HIENTPERDPIIKISAMEIYNEIVKDL--RPESTNRLDLPDPEKGTIVEKEEIAKDS 200
Qy 175 EMALKWTIKGKSRHYGETKMNQSRSHITIFRMILESREKGEPSNCEGSKVSHLNV 234
Db 201 QHLRLHLSICEBQVQVGETALNDTSRSHQIIRLTLESRLR-EVSGCVKSF-VANLNFV 258
Qy 235 LAGSERAQAOTGAAGVRLKEGNCINRSLFILGOVKKLSGQVGGFINYRDSKLTILQNS 294
Db 259 LAGSERAQAOTHAIGARLKEGCHINRSLTLTTVIRKLSGKRSCHIPYRDSKLTILQLS 318
Qy 295 LGGNPKTRIICTITPVSF---DEFTLQALQFASAKYMKNTPYNEV 337
Db 319 LGGNARTAICTMSPAQTHVEQSRNTLFFATCAKEVTNNAKVMV 364
```

```
RESULT 8
US-10-425-114-62672
; Sequence 62672, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 62672
; LENGTH: 955
```

```
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3689-241-All_FLI.pep
; US-10-425-114-62672

Query Match 38.0%; Score 663.5; DB 12; Length 955;
Best Local Similarity 46.8%; Pred. No. 1.1e-58;
Matches 162; Conservative 46; Mismatches 117; Indels 21; Gaps 8;

Qy 1 AEEGAVAVCVVRPLNSREESLGETAQVYWK--TDNNVIYQVDS-----KSNFDRVEH 53
Db 33 AKBEKIFVTVRPLSKKE--LAAKDDVWECADTQTILYKGAQDRAAPMSYTFDKVFG 90
Qy 54 GNETTKNVEIEAIIPIIDSAIOGYNGTIFAYGQTASGKTYTMMGSEDLHGVIPRAIHDF 113
Db 91 PACQTDVVEEGAKDVAMSALTGINATIFAYGQTSSTGKTFTMR-----GVTESAVSDIY 144
Qy 114 QKIKPPDRFLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNRYVADLTVEEVYTS 173
Db 145 RHIDSTPEREFVTKISAMEIYNEIVKDL--RPDSAPFLRLDDPEKGTIVEKLEQETAXD 202
Qy 174 SEMALKWTIKGKSRHYGETKMNQSRSHITIFRMILESREKGEPSNCEGSKVSHLNV 233
Db 203 SOHLRLHLSICEBQVQVGETALNDTSRSHQIIRLTLESRLR-EVSGCVKSF-VANLNFV 260
Qy 234 LAGSERAQAOTGAAGVRLKEGNCINRSLFILGOVKKLSGQVGGFINYRDSKLTILQNS 293
Db 261 LAGSERAQAOTHAIGARLKEGCHINRSLTLTTVIRKLSGKRSCHIPYRDSKLTILQL 320
Qy 294 SLGNPKTRIICTITP--VSFDETLQALQFASAKYMKNTPYNEV 337
Db 321 SLGNARTAICTMSPALTHVEQSRNTLFFATCAKEVTNTAKVMV 366
```

```
RESULT 9
US-10-334-143-33
; Sequence 33, Application US/10334143
; Publication No. US20040009549A1
; GENERAL INFORMATION:
; APPLICANT: GRIGORIEV, IGOR VYACHESLAVOVICH
; APPLICANT: SUDASANAM, SUCHA
; TITLE OF INVENTION: METHOD FOR DETECTING REMOTE HOMOLOGUES AND NOVEL
; TITLE OF INVENTION: KINASES IDENTIFIED WITH THE METHOD
; FILE REFERENCE: 038602/1543
; CURRENT APPLICATION NUMBER: US/10/334,143
; PRIOR FILING DATE: 2002-12-31
; PRIOR FILING DATE: 2001-12-31
; NUMBER OF SEQ ID NOS: 207
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 33
; LENGTH: 1237
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-334-143-33

Query Match 38.0%; Score 663; DB 15; Length 1237;
Best Local Similarity 44.1%; Pred. No. 1.9e-58;
Matches 150; Conservative 54; Mismatches 120; Indels 16; Gaps 6;

Qy 6 VAVCVVRPLNSREESLG-ETAQVYWKTDNNVIYQVDSKSFNDRVPHGNKTYVEE 64
Db 13 VRVALRCRPLVPKEISEGQCMCLSFVPEQTQVVGTD--KSFTYDFVDFPCTEQEEVFNK 70
Qy 65 IAAPIIDSAIOGYNGTIFAYGQTASGKTYTMMG-----SEDLHGVIPRAIHDFOKIK 117
Db 71 AVAPLKGIFKGNATVLAIGQTSKTYSMGAYTAQENEPVGIIPVLIQLLEID 130
Qy 118 KFPDRFLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNRYVADLTVEEVYTS 177
Db 131 KKSDFEFTLAVSYLEIYNEEILDLCPREKAQINIREDPKEGKIVGLTKTVLALDT 190
```

QY 178 LKWTGKSRHYGETKMNORSRSHITIFRMILESEKGEPSNCFNDRVFRHGNETHKNNVYEE 237
Db 191 VSCLEQGNNSRTVASTAMNSQSSRSHTAFTISIEQRKSD-KNCSFR--SKLHLVDLAG 246
QY 238 SERAAQTGAAGVRLKEGCNINRSILFQVVKLSGQVGGFNYRDSKLTTRILQNSLGG 297
Db 247 SERQKTKAEGDRLKEGININRGLLCLGNVISALGDDKGGFAPYRDSKLTTRILQNSLGG 306
QY 298 NPKTRIICITPV--SFDETLTALQFASAKYMNTPYVN 335
Db 307 NSHTLMIACVSPADSNLEETLNTLRYADRARKIKNKPYN 346
RESULT 10
US-10-116-712-664
; Sequence 664, Application US/10116712
; Publication No. US20030194764A1
; GENERAL INFORMATION:
; APPLICANT: Bangor, Chaitanya S.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.568
; CURRENT APPLICATION NUMBER: US/10/116, 712
; CURRENT FILING DATE: 2002-04-07
; NUMBER OF SEQ ID NOS: 670
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 664
; LENGTH: 1232
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-116-712-664

Query Match 37.7%; Score 658; DB 14; Length 1232;
Best Local Similarity 44.7%; Pred. No. 6e-58;
Matches 152; Conservative 52; Mismatches 120; Indels 16; Gaps 6;
QY 6 VAVCVVRPLNSREESLG-ETAQVYWKTDNNVIYQVDSKSFNDRVFRHGNETHKNNVYEE 64
Db 10 VRVALRCRLVPKEISEGCMCLSFVPGEPQVVVGTD--KSFTYDFVDFPSTQEVEFNT 67
QY 65 IAAPIDSALQYNGTIFAYGQTASGKTYTMG-----SEDLGVIPRAIHDFOKIK 117
Db 68 AVAPLKGKGVFNATVLAQGTGSKTYSMGAYTAQENETVGVIPRVQLLPEID 127
QY 118 KFPDREFLLRVSMETINETITDLGCTOKMKPLIREDVNRNVYADLTEEYVYVSEMA 177
Db 128 KKSDFEFLKVSLEYLYNEBEILDLLCPSREKAQINIREDPKEGKIVGLTEKTVLVALDT 187
QY 178 LKWTGKSRHYGETKMNORSRSHITIFRMILESEKGEPSNCFNDRVFRHGNETHKNNVYEE 237
Db 188 VSCLEQGNNSRTVASTAMNSQSSRSHTAFTI---SLEQKSKDNSSFR-SKLHLVDLAG 243
QY 238 SERAAQTGAAGVRLKEGCNINRSILFQVVKLSGQVGGFNYRDSKLTTRILQNSLGG 297
Db 244 SERQKTKAEGDRLKEGININRGLLCLGNVISALGDDKGGFAPYRDSKLTTRILQNSLGG 303
QY 298 NPKTRIICITPV--SFDETLTALQFASAKYMNTPYVN 335
Db 304 NSHTLMIACVSPADSNLEETLNTLRYADRARKIKNKPYN 343

RESULT 11
US-10-116-712-669
; Sequence 669, Application US/10116712
; Publication No. US20030194764A1
; GENERAL INFORMATION:
; APPLICANT: Bangor, Chaitanya S.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.568
; CURRENT APPLICATION NUMBER: US/10/116, 712

; CURRENT FILING DATE: 2002-04-07
; NUMBER OF SEQ ID NOS: 670
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 669
; LENGTH: 1232
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-116-712-669
Query Match 37.7%; Score 658; DB 14; Length 1232;
Best Local Similarity 44.7%; Pred. No. 6e-58;
Matches 152; Conservative 52; Mismatches 120; Indels 16; Gaps 6;
QY 6 VAVCVVRPLNSREESLG-ETAQVYWKTDNNVIYQVDSKSFNDRVFRHGNETHKNNVYEE 64
Db 10 VRVALRCRLVPKEISEGCMCLSFVPGEPQVVVGTD--KSFTYDFVDFPSTQEVEFNT 67
QY 65 IAAPIDSALQYNGTIFAYGQTASGKTYTMG-----SEDLGVIPRAIHDFOKIK 117
Db 68 AVAPLKGKGVFNATVLAQGTGSKTYSMGAYTAQENETVGVIPRVQLLPEID 127
QY 118 KFPDREFLLRVSMETINETITDLGCTOKMKPLIREDVNRNVYADLTEEYVYVSEMA 177
Db 128 KKSDFEFLKVSLEYLYNEBEILDLLCPSREKAQINIREDPKEGKIVGLTEKTVLVALDT 187
QY 178 LKWTGKSRHYGETKMNORSRSHITIFRMILESEKGEPSNCFNDRVFRHGNETHKNNVYEE 237
Db 188 VSCLEQGNNSRTVASTAMNSQSSRSHTAFTI---SLEQKSKDNSSFR-SKLHLVDLAG 243
QY 238 SERAAQTGAAGVRLKEGCNINRSILFQVVKLSGQVGGFNYRDSKLTTRILQNSLGG 297
Db 244 SERQKTKAEGDRLKEGININRGLLCLGNVISALGDDKGGFAPYRDSKLTTRILQNSLGG 303
QY 298 NPKTRIICITPV--SFDETLTALQFASAKYMNTPYVN 335
Db 304 NSHTLMIACVSPADSNLEETLNTLRYADRARKIKNKPYN 343

RESULT 12
US-10-311-642-2
; Sequence 2, Application US/10311642
; Publication No. US2004008687A1
; GENERAL INFORMATION:
; APPLICANT: Cytokinetics, Inc.
; APPLICANT: Beraud, Christophe
; APPLICANT: Freedman, Richard
; TITLE OF INVENTION: NOVEL MOTOR PROTEINS AND METHODS FOR THEIR
; FILE REFERENCE: 020552-001910US
; CURRENT APPLICATION NUMBER: US/10/311,642
; CURRENT FILING DATE: 2003-09-29
; PRIOR APPLICATION NUMBER: 09/597,602
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1029
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: HsKif17 amino acid sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: HsKif17
US-10-311-642-2

Query Match 37.2%; Score 649.5; DB 16; Length 1029;
Best Local Similarity 42.8%; Pred. No. 3.4e-57;
Matches 149; Conservative 57; Mismatches 117; Indels 25; Gaps 8;
QY 5 AVAVCVVRPLNSREESLGETAQVYWKTD-----NNVIYQVDSKSFNDRVFRHGN 56
Db 5 AVKVVVRCRPMQNREREL--RCQPVVTVDCARAQCCIQNPGAADEPPKQFTFDGAYHVDH 62
QY 57 TTKNNVYEEIAAPIIDSAIQYNGTIFAYGQTASGKTYTMGSED--HLGVIPRAIHDF 113

[illegible]

```

RESULT 13
US-10-437-963-114373
; Sequence 114373, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 114373
; LENGTH: 420
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_18070C.1.pep
US-10-437-963-114373

```

Query Match	37.11%	Score 647.5	DB 16	Length 420
Best Local Similarity	43.6%	Pred. NO. 1.4e-57		
Matches 149	Conservative 57	Mismatches 115	Indels 21	Gaps 6
QY	6	VAVCVRVPLNSRESLGHTAQVYWKTDNNVIYQVDS-----KSFNFDRVFHHGNET 57		
DB	17	IVVSVRLPVRNARERGDGD--WECAGPTTLFRGAVPERAMEFPASYSYDVFHSBCG 74		
QY	58	TKNVYEEIAAPIDSIAIOGYNCTIPAYGQTASGKTYTWMGSEDLHGVIPRAIHDFQIK 117		
DB	75	TRQVYDEGARQVAMSVLGINASIPAYGQTSGKTYTVMGITEY-----SMSDIYDIE 128		
QY	118	KPDDREFLLRVSYMEYNETITDLLCGQTKMKPLIREDVNRNVTVADLTSEVVVYSEMA 177		
DB	129	KHPEREFILKFAMEIYNAEVRDLI--SSDATPLRLDDPEKGTVEKLTETLEDKGHL 186		
QY	178	LKWITKGEKSRHYGETKMNQRSRSHITFRMILESREKEGPNCEGVSVKVSHNLVDLAG 237		
DB	187	LELLAVCEAQROIGETAMNEASSRSHQILRMVTVESSAKQFLKGNSSTLIACVNFVDLAG 246		
QY	238	SRAEAQTGAAGVRLKEGCNINSLFILGOVIKKLSDGOGVGFNNVRDSKLTILONSLGG 297		
DB	247	SRAEQTASAGRLKEGSHINSLTLGVIPQLSKGR--NGHIPYRDSKLTILQSSLLG 305		
QY	298	NPKTRIICITTPV--SFDETTLALOPASTAKYMKNTPPYNEV 337		

db 306 NARTAI ICTMSPAHCIEQSRNTLLFANCAKDVTNAQNVV 347

```

RESULT 14
US-10-408-765A-1664
; Sequence 1664, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1664
; LENGTH: 672
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-408-765A-1664

```

Query Match	36.1%	Score 630;	DB 16;	Length 672;
Best Local Similarity	46.8%;	Pred. No. 1.8e-55;		
Matches 145;	Conservative 49;	Mismatches 102;	Indels 14;	Gaps 7;
37	IYVDGS----	KSFNFRVRFHNETTKNVYEEI	AAPIIDSALIGYNGTIFAYGOTASGKT	92
18	VHKTDNNEPPKTF	FTFVFGPESKQLDVYNLTARP	PIIDSVLGYNGTIFAYGOTGIGKT	77
93	YTMGSE----	DHLGVI	PRAIHDIFQKIKFP-DREFLLRVSYMEI	YNETITDILLCGTQM 148
78	FTMEGVRAIPELGI	IPNSFAHIFGHIAKABGDTFLV	RVSYLEIYNEEVRDLL-GKDQT	136
149	KPLIREDNRNVVADL	TEEVVYTSSEMAKWIITKGEKSRHYGETK	KNQRSRSHITFRM 208	
137	QRLEKRPDVGVI	IKOLSAVVANNADMDRIMTLGH	KNRVSGATNMNEHSRSHAITFI 196	
209	ILESEKGEPCSGSVK	YSHNLVDLAGSRAAQ	TGAAGVRLKEGCINRSLFTLGOVI 268	
197	TIECEKIGIDGMH-	-VWNGKLHLVDLAGSRQAQ	TGATGQRLKEATKINLSLTLGNVI 254	
269	KKLSDQGVGGFINVR	DSKLTRILQNSLGNPKTR	IICTITPV--SFDSTLALQPASTAK 326	
255	SALVDGK-STHVPYR	NSKLTRLLQDSLGN	SKTMMCANIGPADVNYDETIST	LYRANRAK 313
327	YMKNTPPYNE	336		
314	NICKARINE	323		

RESULT 15
US-10-437-169150
; Sequence 169150, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: Thomas J.
; APPLICANT: La Rosa, David K.
; APPLICANT: Kovalic, Yihua
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B

```

; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 169150
; LENGTH: 965
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_67597C.1.pap
US-10-437-963-169150

Query Match      35.7%; Score 622.5; DB 16; Length 965;
Best Local Similarity 42.5%; Pred. No. 1.9e-54;
Matches 147; Conservative 53; Mismatches 103; Indels 43; Gaps 7;

QY 1 AEGGAVAVCVVRPLNSRE-----ESLGETAQVYWKTDNNVIYQVDGSKSFNDRVFH 53
Db 17 AKERIMVSVRLPLNGREAGDCDWEICISPTVMFRST---VPERAMFPTATYTDVRFV 73

QY 54 GNETTKNVVEETAAPLIDSAIQYNGTIFAYGQTASGKTYTMGSDHLGVIPRAIHDF 113
Db 74 PDSSTQVYEEGAKVSVSGINSIFAYGQTSSGKTYTWTGITEY-----SVLDIY 127

QY 114 QKIKPPDPREFLLRVSYMEIYNETITDLLCGTQRMKPLIREDVNRNVYVADLTEVVYT 173
Db 128 DYIEKHPEREFILRFSALIEIYNEAVRDLJ--SHDTTFLRLDDPEKGTTVKLTBTETLRD 185

QY 174 SEMALKWITKGEKSHYGETKNQSRSHITFRMILESREKGEPSNCEGSKVSHNLV 233
Db 186 KDHLRNLAVCEAQRQIGETALNETSSRSHQLR-----LNFV 223

QY 234 DIAGSERAAQTGAAGVRLKEGNCINRSLFILGOVIKKLSDGOVGGFINYRDSKLTRELON 293
Db 224 DIAGSERASQTASAGVRLKEGSHINRSLTLTGKVVQKLSKGR--NGHIPRDSKLTRELON 282

QY 294 SLGGNPKTRIICTITPV--SPDETTLAQFASTAKYMKNTPYNYEV 337
Db 283 SLGGNARTAIICTMSPARSHIEQSRNTLLFATCAKEVVTNAQNVV 328

```

Search completed: July 29, 2004, 10:06:25
Job time : 16.9319 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 09:28:21 ; Search time 5.0651 Seconds
(without alignments)
6437.961 Million cell updates/sec

Title: US-10-045-631B-88_COPY_2_340

Perfect score: 1744

Sequence: 1 ABEGAVAVCVVRPLNSRRE.....QFASTAKYMKNTPYVNEVST 339

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1744	100.0	2563	1 S28261	centromere protein
2	1299.5	74.5	2354	2 T14156	kinesin-related pr
3	795.5	45.6	823	2 T52425	kinesin-like prote
4	760	43.6	888	2 D96619	protein T30E16.9 [
5	674	38.6	786	2 A53939	kinesin homolog KH
6	670	38.4	742	1 S58691	kinesin-related pr
7	666.5	38.2	1459	2 T30196	kinesin motor prot
8	665.5	38.2	1231	2 A54803	microtubule-associ
9	663	38.0	701	1 B44259	kinesin-related pr
10	659	37.8	747	1 A57107	kinesin-related pr
11	653.5	37.5	1226	2 T51617	kinesin-like prote
12	643.5	36.9	699	1 S38982	kinesin-related pr
13	633	36.3	672	2 S54351	kinesin oem-3 - Ca
14	633	36.3	932	2 T49235	kinesin-like prote
15	633	36.3	1225	2 A56514	chromokinesin - ch
16	631.5	36.2	909	2 H86350	hypothetical prote
17	621.5	35.6	1121	2 T06065	hypothetical prote
18	620.5	35.6	968	2 T45746	hypothetical prote
19	615.5	35.3	975	1 A31497	kinesin heavy chai
20	614	35.2	793	2 JC5831	kinesin-related pr
21	613	35.1	581	2 B84599	probable kinesin h
22	612.5	35.1	1031	1 A38713	kinesin heavy chai
23	612.5	35.1	1130	2 T21134	hypothetical prote
24	612	35.1	1076	2 B84687	probable kinesin-1
25	611	35.0	1254	2 T18277	kinesin heavy chai
26	609	34.9	1921	2 T13827	kinesin-73 - fruct
27	608.5	34.9	1263	2 T13465	hypothetical prote
28	606.5	34.8	294	2 S38983	kinesin-related pr
29	605	34.7	1066	1 A48669	kinesin-related pr

30	603.5	34.6	554	2 T50118	kinesin-related pr
31	603	34.6	963	1 A41919	kinesin heavy chai
32	602.5	34.5	1032	2 I38510	neuronal kinesin h
33	602.5	34.5	1388	2 T30335	KuP2 protein - Afr
34	598.5	34.3	967	1 A35075	kinesin heavy chai
35	598.5	34.3	1584	1 JN0114	kinesin-related pr
36	598.5	34.3	1584	2 T15822	kinesin-like prote
37	598	34.3	928	2 T10164	kinesin heavy chai
38	597	34.2	332	2 C48835	kinesin-like prote
39	595	34.1	968	2 T51933	kinesin motor prot
40	595	34.1	1027	2 S37711	kinesin heavy chai
41	591	33.9	843	2 S44868	kinesin heavy chai
42	588	33.7	784	1 A55236	kinesin-related pr
43	587.5	33.7	935	2 T51930	kinesin [imported]
44	586.5	33.6	1576	2 T29237	hypothetical prote
45	584	33.5	881	2 I84737	kinesin heavy chai

ALIGNMENTS

RESULT 1

S28261

N/Alternate names: centromere

C/Species: Homo sapiens (man)

C/Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 19-Jan-2001

C/Accession: S28261

R/Yen, T.J.; Li, G.; Schaar, B.T.; Szilak, I.; Cleveland, D.W.

Nature 359, 536-539, 1992

A/Title: CENP-E is a putative kinetochore motor that accumulates just before mitosis.

A/Reference number: S28261; MUID:93024922; PMID:1406971

A/Accession: S28261

A/Molecule type: mRNA

A/Residues: 1-2863 <YEN>

A/Cross-references: EMBL:Z15005; NID:g29864; PIDN:CAA78727.1; PID:g29865

C/Genetics:

A/Gene: GDB:CENPE

A/Cross-references: GDB:361164; OMIM:117143

A/Map position: 4q24-q25

C/Superfamily: centromere protein E; kinesin motor domain homology

C/Keywords: ATP; coiled coil; microtubule binding; mitosis; nucleotide binding; P-loop

F/7-335/Domain: kinesin motor domain homology <KMOT>

F/86-93/Region: nucleotide-binding motif A (P-loop)

F/486-2183/Domain: coiled coil #status predicted <COI>

F/92/Binding site: ATP (lys) #status predicted

Query Match 100.0%; Score 1744; DB 1; Length 2663;
Best Local Similarity 100.0%; Pred. No. 5.6e-124;
Matches 339; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ABEGAVAVCVVRPLNSRRESIGETAQYVYKTDNNVYQVDSKSFNDRVPHGNETTKN	60
DB	2	ABEGAVAVCVVRPLNSRRESIGETAQYVYKTDNNVYQVDSKSFNDRVPHGNETTKN	61
QY	61	VVEETAAPIIDSAIQYNGTIFAYGQTASGKTYTWMSDHLGVIPRAIHDIFQIKIKPP	120
DB	62	VVEETAAPIIDSAIQYNGTIFAYGQTASGKTYTWMSDHLGVIPRAIHDIFQIKIKPP	121
QY	121	DREFLLRVSMEIYNETITDLCGTQMKPLIREDVNRNVYADLTTEVVYVTSMAWK	180
DB	122	DREFLLRVSMEIYNETITDLCGTQMKPLIREDVNRNVYADLTTEVVYVTSMAWK	181
QY	181	ITKGKSRHYGCTKNQSRSHITFRMILSRKGEPSNCEGSKVSHNLVDLAGSR	240
DB	182	ITKGKSRHYGCTKNQSRSHITFRMILSRKGEPSNCEGSKVSHNLVDLAGSR	241
QY	241	AAQTGAAGVRLKEGGINRSLFILGVIKKISDGGVGGFINYRDSKLTILQNSLGGNPK	300
DB	242	AAQTGAAGVRLKEGGINRSLFILGVIKKISDGGVGGFINYRDSKLTILQNSLGGNPK	301
QY	301	TRIICTITPVSFDELTALQFASTAKYMKNTPYVNEVST	339

Db 302 TRIICTITPVSFDETLTALQFASAKYMNTPYVNEVST 340

RESULT 2

T14156

kinesin-related protein - African clawed frog

C:Species: Xenopus laevis (African clawed frog)

C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-May-2000

C:Accession: T14156

R:Wood, K.W.; Sakowicz, R.; Goldstein, L.S.; Cleveland, D.W.

Cell 91, 357-366, 1997

A:Title: KENP-E is a plus end-directed kinetochore motor required for metaphase chromosome segregation

A:Reference number: Z17893; PMID:98028574; PMID:9363944

A:Accession: T14156

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-2954 <WOO>

A:Cross-references: EMBL:AF027728; NID:g2586070; PID:g2586071; PIDN:AAC60300.1

C:Gene: KENP-E

C:Superfamily: centromere protein E; kinesin motor domain homology

Query Match 74.5%; Score 1299.5; DB 2; Length 2954;

Best Local Similarity 73.5%; Pred. No. 4.3e-90;

Matches 250; Conservative 42; Mismatches 43; Indels 5; Gaps 2;

Qy 1 ABEAGAVCVVRPLNSREESLGETAQVYVTKDNNVYQVDSKSFNFDRVFGHNETTKN 60

Db 2 SEGDAVVCVVRPLIQEQ--GDQANLQWAGNTISQVDTGKSFNDRVFNESHESISQ 59

Qy 61 VYEETAAPIDSAIQGYNGTIPAYGOTASGKTYTMGSEDLGLGVTPRAIHDFQKIKFP 120

Db 60 IVORIAVPIIRLSALQGYNGTIPAYGOTSSGKTYTMGTPNSLGIIPQAIQEVFKIIOBIP 119

Qy 121 DREFLLRYSMEIYNETITDLCGTQKMKPLIREDVNRVYVADLTTEEYVYVTSMAKWK 180

Db 120 NREFLLRYSMEIYNETITDLCGTQKMKPLIREDVNRVYVADLTTEEYVYVTSMAKWK 179

Qy 181 ITKGKSRHYGETKNORSSRSHTIFRMILSRKGEPS--NCEGSKVYVSHNLVLDLAG 237

Db 180 IKKGKSRHYGETKNORSSRSHTIFRMILSRKGEPS--NCEGSKVYVSHNLVLDLAG 239

Qy 238 SERAATGAAGVRLKEGNCNINSLFILQGVIKKLSGQGVGGPINTYRDSKLTILQNSLGG 297

Db 240 SERASOTGAAGVRLKEGNCNINSLFILQGVIKKLSGQGVGGPINTYRDSKLTILQNSLGG 299

Qy 298 NPKTILICTITPVSFDETLTALQFASAKYMNTPYVNEV 337

Db 300 NAKTVIICITITPVSFDETLTALQFASAKYMNTPYVNEV 339

RESULT 3

T52425

kinesin-like protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 24-Oct-2000 #sequence_revision 24-Oct-2000 #text_change 24-Oct-2000

C:Accession: T52425

R:Kato, A.; Suzuki, M.; Kuwahara, A.; Ooe, H.; Higano-Inaba, K.; Komeda, Y.

Gene 239, 309-316, 1999

A:Title: Isolation and analysis of cDNA within a 300 kb Arabidopsis thaliana genomic region

A:Reference number: Z55171

A:Accession: T52425

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-823 <KAT>

A:Cross-references: EMBL:AB028468; PIDN:BA88112.1

C:Gene: ZCF125

Query Match 45.6%; Score 795.5; DB 2; Length 823;

Best Local Similarity 51.4%; Pred. No. 1.7e-52;

Matches 178; Conservative 48; Mismatches 95; Indels 25; Gaps 7;

Qy 6 VAVCVVRPLNSREESLGETAQVYVTKDNNVYQVDSKSFN-----FDRVFGHNE 56

Db 4 ICVAVRVRP-----PAPENGASLWKVEDN---RISLHKSLDTPITTASHAFVDFDESS 54

Qy 57 TTKNVYVEIAAPIIDSAIQGYNGTIPAYGOTASGKTYTMGSEDLGLGVTPRAIHDFOKI 116

Db 55 TNASVYELLTWDIIHAIVEGFGNAPAYGQISSKTFNTGSETDGPRIIRSVRDFERI 114

Qy 117 KKFDPDRFLRVSYMEIYNETITDLCGTQKMKPLIREDVNRVYVADLTTEEYVYVTSMA 176

Db 115 HMISDREFLRVSYMEIYNEEINDLL--AVENORLQIHEHLRGVFAVLKEEIVSDAEQ 172

Qy 177 ALKWITGKSRHYGETKNORSSRSHTIFRMILSRKGEPSNCEGSKVYVSHNLVLDLA 236

Db 173 ILKLIDSGEVRNHFGETNMNVHSSRSHTIFRMVIESR--GKDNSSSDAIRVSVLNLVLDLA 230

Qy 237 GSERAAOTGAAGVRLKEGNCNINSLFILQGVIKKLSGQ--QVGGFNTYRDSKLTILQNSL 295

Db 231 GSERIAKTGAGVRLQEGKYINKSLMILGNVINKLSGSTKLAHPTPYEDSKLTILQPAL 290

Qy 296 GGNPKTRIICTITPVSF--FDETLTALQFASAKYMNTPYVNEVST 339

Db 291 GGNAKTCTIICITIAPEEHIEBSKGTLPQFASRAKRTINCAQVNEILT 336

RESULT 4

D96619

protein T30R16.9 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: D96619

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso

Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maity, R.; Marziani

Rizzo, M.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: AB6141; PMID:21016719; PMID:11130712

A:Accession: D96619

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-888 <STO>

A:Cross-references: GB:AB005173; NID:g8778739; PIDN:AAF79747.1; GSPDB:GN00141

C:Gene: T30R16.9

A:Map position: 1

Query Match 43.6%; Score 760; DB 2; Length 888;

Best Local Similarity 46.1%; Pred. No. 9.6e-50;

Matches 178; Conservative 48; Mismatches 98; Indels 62; Gaps 8;

Qy 6 VAVCVVRPLNSREESLGETAQVYVTKDNNVYQVDSKSFN-----FDRVFGHNE 57

Db 4 ICVAVRVRP-----PAPENGASLWKVEDN---RISLHKSLDTPITTASHAFVSGISSTDLI 57

Qy 39 QVDSKSFNF-----DRVFGHNETTKNVYVEIAAPIIDSAIQGYNGTIPAYGQTA 88

Db 58 EIVSLFLSFGVGVYFFLPADHVFDESSTNASVYELLTKDIIHAAVEGNGTAFAYGQTS 117

Qy 89 SGKTYTMGSEDLGLGVTPRAIHDFOKIKKPPDRFLRVSYMEIYNETITDLCGTQK 148

Db 118 SGKTYTMGSETDGPRIIRSVRDFERIHMISDREFLRVSYMEIYNEEINDLL--AVEN 175

Qy 149 KPLIREDVNRVYVADLTTEEYVYVTSMAKWKITGKSRHYGETKNORSSRSHTIFRM 208

Db 176 QRLQIHEHLRGVFAVLKEEIVSDAQILKLIDSGEVRNHFGETNMNVHSSRSHTIFRM 235

Qy 209 -----ILSRKGEPSNCEGSKVYVSHNLVLDLAGSRAAQTAAGVRLKEGNCN 256

236 VRFRSYERDILLVIESR--GKDNSSDAIRVSVNLVLDLGGSRIAKTGAGGVRLQEGKY 293

257 INRSFILQGVKKLSDG-QVGGFINYRDSKLTILQNSLGGNPKTRIICTTPVS--FD 313

294 INKSLMILGNVINKLSDSKTRAHIPYRDSKLTILQALPGNNAKTICTIAPEEHHIE 353

314 EFTLTALQFASAKYMKNTPPYNEVST 339

354 ESKGTLPASRAKRITNCAQVNEILT 379

RESULT 5

A53939

C:Species: Chlamydomonas reinhardtii

C:Date: 07-Oct-1994 #sequence_revision 07-Oct-1994 #text_change 02-Feb-2001

C:Accession: A53939

R:Walther, Z.; Vashishta, M.; Hall, J.L.

J. Cell Biol. 126, 175-188, 1994

A:Title: The Chlamydomonas F1A10 gene encodes a novel kinesin-homologous protein.

A:Reference number: A53939; MUID:94299638; PMID:8027176

A:Accession: A53939

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-786 <WAL>

A:Cross-references: EMBL:U33697; NID:9497696; PIDN:AAA21738.1; PID:9497697

A:Note: authors translated the codon AAC for residue 753 as Asp

C:Genetics:

A:Gene: F1A10

C:Superfamily: unassigned kinesin-related proteins; kinesin motor domain homology

C:Keywords: ATP; coiled coil; nucleotide binding; P-loop

F:11-359/Domain: kinesin motor domain homology <KMT>

F:97-104/Region: nucleotide-binding motif A (P-loop)

Query Match 38.6%; Score 674; DB 2; Length 786;

Best Local Similarity 44.6%; Pred. NO. 2.8e-43;

Matches 158; Conservative 59; Mismatches 113; Indels 24; Gaps 8;

QY 5 AVAVCVVRVPLNSREESLGETAQVYWKTDNNVY---QVDGS---KSFNFDRVFGHNET 57

DB 10 SVKVVRCPLNGKEKADGRSIVDMVDYAGQVKVRPKADASEPPKAFITDQYDWNQC 69

QY 58 TKNVYEETAAPLIDSAIQYNGTIFAYGQTASGKTYTMGSEDH---LGVIPRAIHDFQ 114

DB 70 QRDVFDTARPLIDSCIEGYNGTIFAYGQTGTGKSHMEGKDEPPELRGLIPNTERVFE 129

QY 115 KI-KKFPDREFLLRYSVMEIYNETITDLCGTQWKKELLIPEDVRNVVYADLTTEEVVYT 173

DB 130 ITARDSGTKFEFLVRSYLEIYNEEVRDLL-GKQHSKQMKELKESPDGRVYVKDLSQFVCKN 188

QY 174 SEMALKWITKGSKSHYGETKQNRSSRSHITFRMILESREK-----GEPNSCEGS 224

DB 189 YEEMKVLLAGDKNRQVATLMNQDSSRSHSIFITTEICKLESAAQKPGAKKDSNH 248

QY 225 VKVSHNLNVDLAGSRAAQTGAAGVRLKEGCNNRSUFILQGVIKKLSDGQGVGFINYRD 284

DB 249 VRVGKLNLDLAGSRQDKTGATGRLEKIGIKNLSTALGNVISALVDGK-SGHIPIRD 307

QY 285 SKLTRILQNSLGNPKTRIICTIPV--SFDETALQFASAKYMKNTPPYNE 336

DB 308 SKLTRILODPLSGNTKTMVANGPADWNIDETMTSTLFYANRAKNIQNKPKINE 361

RESULT 6

S58691

N:Alternate names: kinesin-2 chain B; KRP (85/95) 95k chain

C:Species: Strongylocentrotus droebachiensis

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-Jan-2001

C:Accession: S58691

R:Rashad, D.J.; Wedaman, K.P.; Scholey, J.M.

J. Mol. Biol. 252, 157-162, 1995

A:Title: Heterodimerization of the two motor subunits of the heterotrimeric kinesin

Qy 87 TASGKTYPMGSEDLGVIPRAIHDFQIKKFPDREFLLRVSYMEIYNETITDLLCGTQ 146
Db 354 TGSCKTHMTSGDAEPGVIPRAVEQIFQMIKDEPDREFLLRVSYMEIYNETIKOLLA--- 410
Qy 147 KKKPL-----IREDVNRVYVADLTVV 171
Db 411 PLPLPTGSGSLQTTDRPASPIKGGSSHAAGSQSCCTLRIEDQKSRVITGLUREIV 470
Qy 172 YTSEMALKWITKGEKSRHVGTEKMNORSRSHITFRMILESEKGEPCNCEGSKVSHLN 231
Db 471 TDANTVLCIORQDERHVGAIDWNNERSRSHCVFQLIERSPAPSASKE--VRISQIN 528
Qy 232 IVDLAGSRAAQTAGVRLKRGECNINRSLFILGVQIKKLDGQVGG--FINYRDSKLT 289
Db 529 LIDLAGSRAA---SOAERRKEGAFINKSLTLGTIVGKLTPEVNGDAHIPYRDSKLT 585
Qy 290 ILQNSLGNPKTRICTITPV--SFDETLTALQFASTAKYMKNT 331
Db 586 ILQTSLSGNARIIVICTLSPDTEHANETLSTLKFGKCKLVVTT 629
RESULT 8
A54803
Microtubule-associated motor KIF4 - mouse
N:Alternate names: kinesin-related protein KIF4
C:Species: Mus musculus (house mouse)
C>Date: 05-Apr-1995 #sequence_revision 05-Apr-1995 #text_change 02-Feb-2001
C:Accession: A54803; D44259
R:Aizawa, H.; Sekine, Y.; Noda, Y.; Kondo, S.; Aizawa, H.; Takemura, R.; Hirokawa, N.
J. Cell Biol. 127, 187-201, 1994
A:Title: A novel microtubule-based motor protein (KIF4) for organelle transports, whose
F:369-599/Domain: helical rod #status predicted <RDP>
A:Reference number: A54803; MUID:95014709; PMID:7923562
C:Accession: A54803
A:Molecule type: mRNA
A:Residues: 1-1231 <SEK>
A:Cross-references: GB:D12646; NID:9563772; PIDN:BA02167.1; PID:d102657; PID:g563773
R:Aizawa, H.; Sekine, Y.; Takemura, R.; Zhang, Z.; Nangaku, M.; Hirokawa, N.
J. Cell Biol. 119, 1287-1296, 1992
A:Title: Kinesin family in murine central nervous system.
A:Reference number: A44259; MUID:93077686; PMID:1447303
A:Accession: D44259
A:Status: nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 91-111, 'S', 113-240 <AIZ>
A:Experimental source: brain
A:Note: sequence extracted from NCBI backbone (NCBIP:118904)
C:Superfamily: unassigned kinesin-related proteins; kinesin motor domain homology
C:Keywords: ATP; nucleotide binding; P-loop
F:10-343/Domain: kinesin motor domain homology <KNOT>
F:88-95/Region: nucleotide-binding motif A (P-loop)

Query Match 38.2%; Score 665.5; DB 2; Length 1231;
Best Local Similarity 44.3%; Pred. No. 2.3e-42;
Matches 151; Conservative 55; Mismatches 118; Indels 17; Gaps 7;
Qy 6 VAVCVVRPLNSREESLG-ETAOVYWKTDNNVYQVDSGKSFNDFRVFHNKNTKNVYEE 64
Db 10 VRVALRCRPLVSKETKEGCQTCLSFVGPGEQVW--VGNDKSFTYDFVDPDPSTQEVEPNT 67
Qy 65 IAAPIIDSAIQYNGTIFAYGQTASGKTYTMGWS-----EDH--LGVIPRAIHDFQIK 117
Db 68 AVAPLIKGVFGYNATVLAIGTSGKTYSMGAYTAQOEHDSATGIVIPRIQLLFKEIN 127
Qy 118 KFPDREFLLRVSYMEIYNETITDLLCGT-QRMKPLIREDVNRVYVADLTVVYVYSEM 176
Db 128 KKSDFEFTLVKSYLEIYNEEILDLLCSSRERATQINIREDPKEGKIIVGLETKTVLVASD 187
Qy 177 ALKWKITKGEKSRHVGTEKMNORSRSHITFRMILESEKGEPCNCEGSKVSHNLVDLA 236
Db 188 TVSCLEQNGNRTVASTAMNQSQRSHAFITISIEQRKK-----NDKNSFRSKLHLVDLA 243
Qy 237 GSERAAQTAGVRLKRGECNINRSLFILGVQIKKLDGQVGGFINYRDSKLTTRILQNSLG 296

Db 244 GSERQKTKABGDRLREGININRGLLCGNVISALGDDKGNFVYRDSKLTTRLLQDSL 303
Qy 297 GNPKTRIICTITPV--SFDETLTALQFASTAKYMKNTPVVN 335
Db 304 GNSHTLACVSPADSNLEETLNTLRYADRARKINKKEPIIN 344
RESULT 9
B44259
Kinesin-related protein KIF3A - mouse
C:Species: Mus musculus (house mouse)
C>Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 19-Jan-2001
C:Accession: B44259; S27872
R:Aizawa, H.; Sekine, Y.; Takemura, R.; Zhang, Z.; Nangaku, M.; Hirokawa, N.
J. Cell Biol. 119, 1287-1296, 1992
A:Title: Kinesin family in murine central nervous system.
A:Reference number: A44259; MUID:93077686; PMID:1447303
A:Accession: B44259
A:Molecule type: mRNA
A:Residues: 1-701 <AIZ>
A:Experimental source: brain
A:Cross-references: EMBL:D12645; NID:G220469; PIDN:BA02166.1; PID:g220470
A:Note: sequence extracted from NCBI backbone (NCBIP:118911)
C:Complex: heterodimer with KIF3B (PIR:A57107); the KIF3A/3B heterodimer associates with
C:Function:
A:Description: KIF3 complex is a motor protein that provides anterograde fast axonal tr
A:Superfamily: kinesin-related protein KIF3; kinesin motor domain homology
C:Keywords: ATP; coiled coil; heterodimer; heterotrimer; microtubule binding; nucleotid
F:1-368/Domain: head globular #status predicted <HGI>
F:15-354/Domain: kinesin motor domain homology <KNOT>
F:100-107/Region: nucleotide-binding motif A (P-loop)
F:369-599/Domain: helical rod #status predicted <RDP>
F:600-701/Domain: tail globular #status predicted <TGL>
F:106/Binding site: ATP (Lys) #status predicted

Query Match 38.0%; Score 663; DB 1; Length 701;
Best Local Similarity 46.1%; Pred. No. 1.7e-42;
Matches 158; Conservative 50; Mismatches 119; Indels 16; Gaps 8;
Qy 6 VAVCVVRPLNSREESLG-GETAOVYWKTDNNVYQVDSG----KSFNDFRVFHNKNTTK 59
Db 15 VKVVRCPPLNERKSCMYRAQSVDEMGRGITVHKTDSSNEPPKFTTFDVFVGESQL 74
Qy 60 NVYBEIAAPLIDSAIQYNGTIFAYGQTASGKTYTMGSE---DHLGVIPRAIHDFQIK 116
Db 75 DVYNLTARPIIDSVLEGYNGTIFAYGQTGKTFTMEGVRAVPGRLRGVIPNSFAHIFGI 134
Qy 117 KKFP-DREFLLRVSYMEIYNETITDLLCGTQRMKPLIREDVNRVYVADLTVVYVYSE 175
Db 135 AKAEGRTRFLRVSVLEIYNEEVRDLL-GKDQTORLEVKRPDGVYIKDLSAYVYNNAD 193
Qy 176 MALKWITKGEKSRHVGTEKMNORSRSHITFRMILESEKGEPCNCEGSKVSHNLVDL 235
Db 194 DMDRIMTLGHKNSRVGATNNHESRSHAFITIECKEKGVDGNMH--VRMGKHLVDL 251
Qy 236 AGSERRAAQTAGVRLKRGECNINRSLFILGVQIKKLDGQVGGFINYRDSKLTTRILQNSL 295
Db 252 AGSERQAKTAGTQRLKEATKINLSLTGNVISALVDGK-STHVPYRNSKLTTRLLQDSL 310
Qy 296 GGNPKTRIICTITPV--SFDETLTALQFASTAKYMKNTPVYNE 336
Db 311 GGNKTMWCANIGPADYNYDETISTLRYANRAKNKKNKARINE 353
RESULT 10
A57107
Kinesin-related protein KIF3B - mouse
C:Species: Mus musculus (house mouse)
C>Date: 03-Nov-1995 #sequence_revision 03-Nov-1995 #text_change 19-Jan-2001
C:Accession: A57107
R:Yamazaki, H.; Nakata, T.; Okada, Y.; Hirokawa, N.
J. Cell Biol. 130, 1387-1399, 1995
A:Title: KIF3A/B: a heterodimeric kinesin superfamily protein that works as a microtubu


```

112 IFQKIYKKFFDR-EFLILRVSMELVNETITDILLCGTQXMKPLIIREDVNRNVVADLTTEV 170
127 IFGHIAKEQENVRFVRVSYLETYNBEVKOLL-GKQOQHELVKVRPDVGYYVKDL$AFV 185
171 VYTSMALKWITKGEKSRHVGTEKMNORSRSHITFRMILESREKGEPSNCEGSKVQVSHL 230
186 VNNADDMDRIMTILGNKNRSVGATWNNESSRSHAFITILERSDMG--LQKEQHVVRVCKL 243
231 NLVDDLAGEPAAQTGAAGVRLKGCNNINRSLFILGQVVKKLSGQGVGFINYRDSKLTTRI 290
244 HMVDLAGSERQTKTGATGQRGLKEATKINLSILSTLGNVSISSLVGDK-STHPIYRNSKLTRL 302
291 LQNSLGNPKTRITICTITPV--SPDETILTAQFASTAKYMKNTPYNNE 336
303 LQDSLGNNAKTVMCANIGPAEYNYDETISTLRVANRAKNIKKNKAKINE 350

RESULT 13
S54351
kinesin osm-3 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 02-Feb-2001
C/Accession: S54351
R/Tabish, M.; Siddiqui, Z.K.; Nishikawa, K.; Siddiqui, S.S.
J. Mol. Biol. 247, 377-389, 1995
A/Title: Exclusive expression of C. elegans osm-3 kinesin gene in chemosensory neurons
A/Reference number: S54351; MUID:95230679; PMID:7714894
A/Accession: S54351
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-672 <TAB>
A/Cross-references: GB:D38632; NID:g808864; PIDN:BA07612.1; PID:dl008194; PID:gl490195
C/Suprafamily: unassigned kinesin-related proteins; kinesin motor domain homology
C/Keywords: ATP; nucleotide binding; P-loop
F/48-376/Domain: kinesin motor domain homology <KMOT>
F/430-137/Region: nucleotide-binding motif A (P-loop)

```

Query Match	36.3%;	Score 633;	DB 2;	Length 672;
Best Local Similarity	47.0%;	Pred. No. 3e-40;		
Matches	142;	Conservative 49;	Mismatches 97;	Indels 14; Gaps 6;
QY	41	DG-SKSFNFRVPHGNETTKNVVEEIAAPIIDSALQGVNGTIFAYGQTASGKVTMMGSE	99	
DB	85	DGAAKDTFCAGYPMIRPGQIYNDIVFPLVENVIEGTGTVFAYGQTSGKTFSMOGIE	144	
QY	100	D---HLGVIIPRAIHDFQIKKKFPDPREFLLRVSYMEIYNETITDLLCGCTOKMKPLIIRE	156	
DB	145	TIPAQRGVIPAPDPIHTATATTENVKFLVHCSYLEIYNEEVRDLLGADNKQK-LEIQE	203	
QY	157	VNRNVVADTETEYVVTSEMAWKMTYGEKSRHYGETGMQRQSRSTITFRMILESREK	216	
DB	204	PDRGVYVAGLSMHVCHDVPVAPKELMTGFRNNRHVGATLMNKDSRSHSIFTVYVEGITE-	262	
QY	217	EPSNCEGSVKVSHLNLYDLAGSRAAOTGAAGVRLKEGCNINRSLFILGOVKKLSGOQV	276	
DB	263	-----TGSIRMGKUNLYDLAGSERQSKTGATGRLKEATKINLSALGNVISALVDGK-	316	
QY	277	GGFINYRDSKLTIRLQNSLGNPNKTRITICITPVSY--FDETLTALQFASAKYMKNTPYV	334	
DB	317	SKHIPYRDSKLTIRLQDSLGNTKTIACVSPSSDNYDETLSTLRANRAKNIKPKPTI	376	
QY	335	NE 336		
DB	377	NE 378		

```

RESULT 14
T49235
kinesin-like protein - Arabidopsis thaliana
N;Alternate names: protein F7K15.60
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
C;Accession: T49235

```

R;Obermaier, B.; Ottenwaelder, B.; Duchemin, D.; Zeitler, K.; Mewes, H.W.; Rudd, S.; submitted to the Protein Sequence Database, April 2000
A;Reference number: Z25019
A;Accession: T49235
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-932 <OBE>
A;Cross-references: EMBL:AL353871; GSPDB:GN00061; ATSP:F7KJ5.60
A;Experimental source: cultivar Columbia; BAC clone F7K15
C;Genetics:
A;Gene: ATSP:F7KJ5.60
A;Map position: 3
A;Intons: 76/1; 106/1; 139/3; 177/3; 205/1; 227/3; 247/3; 289/2; 355/3; 401/3; 430/3;

Query Match	36.3%;	Score 633;	DB 2;	Length 932;
Best Local Similarity	44.3%;	Prod. No. 4.7e-40;		
Matches 152;	Conservative 55;	Mismatches 116;	Indels 20;	Gaps 8;
Qy	3	EGAVAVCVRPVPLNSREESLGETAQVYWK--TDNNVIQ-----VDGSKSFNFDRVFHGNE	56	
Db	26	EKILVTVRPVLNRRHAKYDL--IAMECPDDETIIVFKNPENPKATKYSFDFKVFPEPTC	83	
Qy	57	TTKNVVEBIAAPIIDSAIOQNGYTIIPAYGQTASGKTYTMGSEDLGLVPIRAIHDFQKI	116	
Db	84	ATQEVYEGGSRDVALSAGTNATIPAYGQTSSGKTFTMR-----GVTESVVKDIYEHI	137	
Qy	117	KKPPDRFLRVSMETIYNETITDLCGTQMKPLIREDVNRNVYADITEEVVYVYSEM	176	
Db	138	RKTQERSFVLKVSALETIYNETVVDLL--NRDTGFLRLDDDFEKTIVENIVEEVYVESRQH	195	
Qy	177	ALKWITKGEKSRHVGETKMNORSRSHITPRMILLESKEGPSCNCEGVKSHLNLVDLA	236	
Db	196	LQHLISICEPQOVGETALMDKSRSHQIIRLTIHSLR--ETAGCVQSF-MATLNLVDLA	253	
Qy	237	GSRAAQGTGAAGVRLKEGCNNINSLFILGQVKKLSDGQGVGFINYRDSKLTRILQNSLG	296	
Db	254	GSERAFQTNADGLRLKEGSHINSELLTITVIRKLSGRKRDRHVPYRDSKLTRILQNSLG	313	
Qy	297	GNPKTRILICTITPV--SFDETLTALQAFSTAKYMKNTPIYNEV	337	
Db	314	GNARTALICTISPAISLHVETQKTKLISFAMSAKEVNCANVMV	356	

RESULT 15
A56514
Chromokinesin - chicken
C:Species: Gallus gallus (chicken)
C:Date: 21-Jul-1995 #sequence_revision 28-Jul-1995 #text_change 02-Feb-2001
C:Accession: A56514; I50691
R:Wang, S.Z.; Adler, R.
J. Cell Biol. 128, 761-768, 1995
A:Title: Chromokinesin: a DNA-binding, kinesin-like nuclear protein.
A:Reference number: A56514; MUID:95181533; PMID:7876303
A:Accession: A56514
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1225 <WAN>
A:Cross-references: GB:U18309; NID:g603760; PIDN:AAC59666.1; PID:g603761
R:Wang, S.Z.; Adler, R. U.S.A. 91, 1351-1355, 1994
Proc. Natl. Acad. Sci.
A:Title: A developmentally regulated basic-leucine zipper-like gene and its expression
A:Reference number: A53451; MUID:94151328; PMID:8108415
A:Accession: I50691
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 728-1086, 'RI' <WA2>
A:Cross-references: EMBL:U04821; NID:g440792; PIDN:AA18960.1; PID:g440793
C:Genetics:
A:Gene: sw3-3
C:Superfamily: unassigned kinesin-related proteins; kinesin motor domain homology
C:Keywords: ATP; DNA binding; mitosis; nucleotide binding; P-loop
F:11-344/Domain: kinesin motor domain homology <NMOT>
F:88-95/Region: nucleotide-binding motif A (P-loop)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 09:26:36 ; Search time 3.25614 Seconds
(without alignments)
5421.082 Million cell updates/sec

Title: US-10-045-631B-88_COPY_2_340

Perfect score: 1744

Sequence: 1 AEEGAVAVCVRRVRLNLSREE.....QFASTAKYKNTFYNEVST 339

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1744	100.0	2663	1	CENE_HUMAN	Q02224 homo sapien
2	674	38.6	786	1	FL10_CHLRE	P46869 chlamydomon
3	670	38.4	742	1	K121_STRPU	P46871 strongyloce
4	668	38.3	1232	1	KF4A_HUMAN	O95239 homo sapien
5	665.5	38.2	1231	1	KF4A_MOUSE	P33174 mus musculu
6	664	38.1	702	1	KF3A_HUMAN	Q9Y496 homo sapien
7	663	38.0	701	1	KF3A_MOUSE	P28741 mus musculu
8	660	37.8	747	1	KF3B_HUMAN	O15066 homo sapien
9	659	37.8	747	1	KF3B_MOUSE	Q61771 mus musculu
10	653.5	37.5	1226	1	KF4A_XENLA	Q91784 xenopus lae
11	649.5	37.2	1029	1	KF17_HUMAN	O9p2e2 homo sapien
12	643.5	36.9	699	1	K122_STRPU	P46872 strongyloce
13	641.5	36.8	1038	1	KF17_MOUSE	Q99pw8 mus musculu
14	633	36.3	672	1	OSM3_CAEEL	P46873 caenorhabdi
15	633	36.3	1225	1	KF4A_CHICK	Q90640 gallus gall
16	615.5	35.3	975	1	KINH_DROME	P17210 drosophila
17	614	35.2	793	1	KF3C_HUMAN	O14782 homo sapien
18	613	35.1	796	1	KF3C_RAT	O55165 rattus norv
19	612.5	35.1	1031	1	KINH_STRPU	P35978 strongyloce
20	611	35.0	796	1	KF3C_MOUSE	O35066 mus musculu
21	605	34.7	1066	1	KL61_DROME	P46863 drosophila
22	604	34.6	963	1	KINH_MOUSE	O9us60 schizosacch
23	603.5	34.6	554	1	KLP3_SCHPO	P33176 homo sapien
24	603	34.6	963	1	KINH_HUMAN	Q12840 homo sapien
25	602.5	34.5	1032	1	KINN_HUMAN	Q9eqw7 mus musculu
26	602	34.5	1749	1	K13A_MOUSE	P33175 mus musculu
27	600.5	34.4	1027	1	KINN_MOUSE	O60282 homo sapien
28	599	34.3	957	1	KF5C_HUMAN	P21613 loligo peal
29	598.5	34.3	967	1	KINH_LOLPE	P23678 caenorhabdi
30	598.5	34.3	1584	1	U104_CAEEL	P48467 neurospora
31	598	34.3	928	1	KINH_NEUCR	Q9nh9 homo sapien
32	598	34.3	1805	1	K13A_HUMAN	Q9nqt8 homo sapien
33	595.5	34.1	1826	1	K13B_HUMAN	

ALIGNMENTS

RESULT 1

ID	CENE_HUMAN	STANDARD;	PRT;	2663 AA.
AC	Q02224;			
DT	01-JUL-1993 (Rel. 26, Created)			
DT	01-JUL-1993 (Rel. 26, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Centromeric protein E (CENP-E protein).			
GN	CENPE			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=93024922; PubMed=1406971;			
RA	Yen T.J., Li G., Schaar B.T., Szilak I., Cleveland D.W.;			
RT	"CENP-E is a putative kinetochore motor that accumulates just before			
RT	mitosis."			
RL	Nature 359:536-539 (1992).			
RN	[2]			
RP	CHARACTERIZATION.			
RX	MEDLINE=95196755; PubMed=7889940;			
RA	Thrower D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;			
RT	"Mitotic HeLa cells contain a CENP-E-associated minus end-directed			
RT	microtubule motor."			
RL	EMBO J. 14:918-926 (1995).			
RN	[3]			
RP	CHARACTERIZATION.			
RX	MEDLINE=98437347; PubMed=9763420;			
RA	Chan G.K.T., Schaar B.T., Yen T.J.;			
RT	"Characterization of the kinetochore binding domain of CENP-E reveals			
RT	interactions with the kinetochore proteins CENP-F and hBUBR1."			
RN	[4]			
RP	FARNESYLATION.			
RX	MEDLINE=20459117; PubMed=10852915;			
RA	Ashar H.R., James L., Gray K., Carr D., Black S., Armstrong L.,			
RA	Bishop W.R., Kirschmeier P.;			
RT	"Farnesyl transferase inhibitors block the farnesylation of CENP-E			
RT	and CENP-F and alter the association of CENP-E with the			
RT	microtubules."			
RN	J. Biol. Chem. 275:30451-30457 (2000).			
CC	-!- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE			
CC	KINETOCHORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE			
CC	OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT			
CC	AND/OR SPINDLE ELONGATION.			
CC	-!- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.			
CC	-!- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCHORES DURING			
CC	CONGRESSION. RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS			
CC	QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.			
CC	-!- SIMILARITY: Belongs to the kinesin-like protein family.			

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/>) or send an email to license@isb-sib.ch.

EMBL; Z15005; CAA78727.1; -.
PIR; S28261; S28261.
HSP; P17119; 3KAR.
Genew; HGNC:1856; CENPE.
GK; Q02224; -.
MIM; 117143; -.
GO; GO:0005699; C:kinetochore; TAS.
GO; GO:0005634; C:nucleus; TAS.
GO; GO:0008350; F:kinetochore motor activity; TAS.
GO; GO:0000067; P:DNA replication and chromosome cycle; TAS.
GO; GO:0007079; P:mitotic chromosome movement; TAS.
GO; GO:0007080; P:mitotic metaphase plate congression; TAS.
InterPro; IPR001752; Kinesin_motor.
Pfam; PF00225; kinesin; 1.
PRINTS; PR00380; KINESINHEAVY.
SMART; SM00129; KISC; 1.
PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
PROSITE; PS00067; KINESIN MOTOR DOMAIN2; 1.
Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis; Cell cycle; Centromere; Lipoprotein; Prenylation.
KINESIN-MOTOR.
DOMAIN 1 335
DOMAIN 336 2471
DOMAIN 2472 2663
NP_BIND 86 93
LIPID 2660 2660
SEQUENCE 2663 AA; 312087 MW; CEFCL3880C8C8CB8 CRC64;

Query Match 100.0%; Score 1744; DB 1; Length 2663;
Best Local Similarity 100.0%; Pred. No. 9.3e-125;
Matches 339; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 AEGAGAVCVVRPLNRSREESLGETAQYWKTDNNVIYQVDGSKSFNDRVPHGNETTKN 60
2 AEGAGAVCVVRPLNRSREESLGETAQYWKTDNNVIYQVDGSKSFNDRVPHGNETTKN 61
61 VYIEIAPIIDSAIQGYNGTIFAYCOTASGKTYTMGSEDLHGVIPRAIHDFQIKKFP 120
62 VYIEIAPIIDSAIQGYNGTIFAYCOTASGKTYTMGSEDLHGVIPRAIHDFQIKKFP 121
121 DRELLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNYYVADLTEEVYVTSSEALKW 180
122 DRELLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNYYVADLTEEVYVTSSEALKW 181
181 ITGKESRHYGETKQNRSSRSHTIFRMILSRKGEPSNCEGSKVSHLNLDLAGSER 240
182 ITGKESRHYGETKQNRSSRSHTIFRMILSRKGEPSNCEGSKVSHLNLDLAGSER 241
241 AAO7GAAGVRLKEGNCINRSFILGOVTKLSDGQGVGFVNRDLSKLTRILQNSLGGNPK 300
242 AAO7GAAGVRLKEGNCINRSFILGOVTKLSDGQGVGFVNRDLSKLTRILQNSLGGNPK 301
301 TRIICITVPSPDETLTALQFASTAKYMKNTPYNEVST 339
302 TRIICITVPSPDETLTALQFASTAKYMKNTPYNEVST 340

RESULT 2
FL10 CHLRE STANDARD; PRT; 786 AA.
AC P46869;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Kinesin-like protein FLA10 (KHP1 protein).
GN FLA10.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.

NCBI_TaxID=3055;
[1]
SEQUENCE FROM N.A.
RC STRAIN=137;
RX MEDLINE=94299638; PubMed=8027176;
RA Walther Z., Vashishtha M., Hall J.L.;
RT "The Chlamydomonas FLA10 gene encodes a novel kinesin-homologous protein.";
RL J. Cell Biol. 126:175-188(1994).
CC -!- FUNCTION: Probably involved in flagellar assembly and maintenance.
CC May play a role in flagellar synthesis.
CC -!- TISSUE SPECIFICITY: Flagellar axoneme.
CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN II SUBFAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/>) or send an email to license@isb-sib.ch.

EMBL; L33697; AAA21738.1; -.
PIR; A53939; A53939.
HSP; P17119; 3KAR.
InterPro; IPR001752; kinesin_motor.
Pfam; PF00225; kinesin; 1.
PRINTS; PR00380; KINESINHEAVY.
SMART; SM00129; KISC; 1.
PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
PROSITE; PS00067; KINESIN MOTOR DOMAIN2; 1.
Motor protein; Microtubule; ATP-binding; Coiled coil.
KINESIN-MOTOR (BY SIMILARITY).
DOMAIN 1 358
DOMAIN 367 687
DOMAIN 688 786
NP_BIND 97 104
POLY-GLY. 391
DOMAIN 388 714
POLY-GLY. 714
DOMAIN 756 759
POLY-ASP.
SEQUENCE 786 AA; 86671 MW; F90969203EB79F1B CRC64;

Query Match 38.6%; Score 674; DB 1; Length 786;
Best Local Similarity 44.6%; Pred. No. 9e-44;
Matches 158; Conservative 59; Mismatches 113; Indels 24; Gaps 8;

5 AVAYCVVRPLNRSREESLGETAQYWKTDNNVIYQVDGSKSFNDRVPHGNET 57
10 SVKVVVRCPLNGKEKADGRSRIVDMVDAGQVKNRPKADASEPPKAFTEQVYDMNCQ 69
58 TQNYIEIAPIIDSAIQGYNGTIFAYCOTASGKTYTMGSEDLHGVIPRAIHDFQ 114
70 QRDVFDITAREPLIDSCIEGYNGTIFAYCOTASGKTYTMGSEDLHGVIPRAIHDFQ 129
115 KI-VKFPDREPLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNYYVADLTEEVYV 173
130 IIAADSQTKELVRSVLEIYNEEVRLD-L-GKDSKMKELKESPRGVYVVDLSQFVCKN 188
174 SEMALKWITTKGKSRHYGETKQNRSSRSHTIFRMILSRKGEPSNCEGSKVSHLNLDLAGSER 224
189 YEEMNKVLLAGKDNQVGTATLMNQDSSRSHTIFRMILSRKGEPSNCEGSKVSHLNLDLAGSER 248
225 VYVSHLNLDLAGSERAAQTGAAGVRLKEGNCINRSFILGOVTKLSDGQGVGFVNRDLSKLTRILQNSLGGNPK 284
249 VRVGKLNLDLAGSERQDKTGATGURLKEGKINLSLTALGNVISALVDGK-SGHIPYRD 307
285 SKLTRILQNSLGGNPKTRIICTITPV--SFDETLTALQFASTAKYMKNTPYNE 336
308 SKLTRILQNSLGGNPKTRIICTITPV--SFDETLTALQFASTAKYMKNTPYNE 361

RESULT 3
KI21_STRPU


```
DR InterPro; IPR001752; kinesin_motor.
DR Pfam; PF00225; kinesin: 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; Kisc; 1.
DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
DR PROSITE; PS0067; KINESIN MOTOR DOMAIN2; 1.
DR Motor protein; Microtubule; ATP-binding; DNA-binding;
KW Nuclear protein; Coiled coil.
FT DOMAIN 1 349 KINESIN-MOTOR.
FT DOMAIN 350 999 COILED COIL (BY SIMILARITY).
FT DOMAIN 1000 1232 GLOBULAR.
FT NP BIND 88 95 ATP (POTENTIAL).
FT CONFLICT 223 223 R -> G (IN REF. 2).
FT CONFLICT 231 231 S -> T (IN REF. 4).
FT CONFLICT 286 286 V -> A (IN REF. 2).
FT CONFLICT 422 422 L -> W (IN REF. 2).
FT CONFLICT 564 564 L -> H (IN REF. 4).
FT CONFLICT 564 564 L -> P (IN REF. 2).
FT CONFLICT 600 600 K -> E (IN REF. 3).
FT CONFLICT 668 668 R -> K (IN REF. 1).
FT CONFLICT 928 928 Q -> P (IN REF. 1).
FT CONFLICT 958 958 Q -> R (IN REF. 3).
FT CONFLICT 960 960 L -> Q (IN REF. 1).
FT CONFLICT 960 960 LL -> S (IN REF. 4).
FT CONFLICT 996 997 QKHPKDTLLSP -> RFLPRIPFYQ (IN REF. 4).
FT CONFLICT 1003 1014 P -> Q (IN REF. 2).
FT CONFLICT 1022 1022 K -> N (IN REF. 2).
FT CONFLICT 1077 1077 K -> N (IN REF. 2).
FT CONFLICT 1138 1138 G -> S (IN REF. 2).
FT CONFLICT 1232 1232 AA; 139908 MW; FF74052A17A88F7 CRC64;
SQ SEQUENCE 1232 AA; 139908 MW; FF74052A17A88F7 CRC64;

Query Match 38.3%; Score 668; DB 1; Length 1232;
Best Local Similarity 44.3%; Pred. No. 4.8e-43;
Matches 152; Conservative 52; Mismatches 120; Indels 16; Gaps 5;

QY 6 VAVCVVRPLNSREESLG-ETAQVYWKTDNNVIYQVDSKSFNDRVPHGNETTKNVEE 64
DB 10 VRVALRCRPLVSEISGQCCLSFVPGEQVGTD--KSFTYDFVDFDSTEQEEVNT 67
QY 65 IAAPLIDSAIQYNGTIFAYQOTASGKTYTMWG-----SEDLGVIPRAIHDFOKIK 117
DB 68 AVAPLIRKGVFGYNATVLAGYQGTSGKTYSGGAYTAEQENPTVGIVPRVQLLFXEID 127
QY 118 KFPDFEFLRVSMEIYNETITDLCGTQKMKPLIIREDVNRNVYVADLFEVVTSEMA 177
DB 128 KKSDFEFTLVKVSYLEIYNEELDLDCPSREKAQINIREDPKGIKIVGLTEKTVLALDT 187
QY 178 LKWTGKESRHYGETKQNRSSHTTIFRMILSRKGEPSNCEGSKVSHLNLVDLAG 237
DB 188 VSCLEQGNNSRTVASTAMNSQSRSHAFTISLQRKSD----KNSSFRSKHLVDLAG 243
QY 238 SERAAQTGAAGVRLKEGGINRSLFILQGVIKLSDGVGGINFYRDSKLTIRLQNSLGG 297
DB 244 SERQKTKAEGDRLKEGININGLLCIGNVLSALGDDKKGFGVYRDSKLTIRLQNSLGG 303
QY 298 NPKTRIITCTPPV--SFDETILTAQFASTAKYKNTPVYN 335
DB 304 NSHTLMIACVSPADSNLETLNTRYADRAKIKNKPIVN 343

RESULT 5
KF4A MOUSE STANDARD; PRT; 1231 AA.
AC P33174;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chromosome-associated kinesin KIF4A (Chromokinesin).
GN KIF4A OR KIF4 OR KNS4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
```

```

QY 237 GSERAQOTGAAGVRLKEGNCINRSFILGQVKKLSDGVGFNVRDCKLTRLQNSLG 296
D 244 GSERQKTKAEGDLREGINRGLCLGNVISALGDDKGNFVPVRDCKLTRLQDQSLG 303
QY 297 GNPKTRIICTITPV--SFDETLTALQFASATKYMKNTPYV 335
D 304 GNSHTLMIACVSPADSNLEETLNTLRYADRAKIKNKPIIN 344

RESULT 6
KF3A HUMAN
ID KF3A HUMAN STANDARD; PRT; 702 AA.
AC Q9Y496; Q86XE9; Q9Y6V4;
DC 30-MAY-2000 (Rel. 39, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Kinesin-like protein KIF3A (Microtubule plus end-directed kinesin motor 3A).
GN KIF3A OR KIF3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Retina;
RX MEDLINE=20018159; PubMed=10548469;
RA Whitehead J.L., Wang S.Y., Bost-Usinger L., Hoang E., Frazer K.A., Burnside B.;
RT "Photoreceptor localization of the KIF3A and KIF3B subunits of the heterotrimeric microtubule motor kinesin II in vertebrate retina."; Exp. Eye Res. 69:491-503 (1999).
RL [2]
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=22386257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marushina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Buterfield V.S.N., Krzywinski M.I., Skalska U., Smalusz D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RA "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences."; Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RL [3]
RP SEQUENCE FROM N.A.
RA Connolly K.S., Gunning K.M., Davis C.A., Kadner K., Subramanian S., Miguel T., Lewis K.D., Fridlyand J., Alcivar D., Benke J.A., Bondoc M., Bowen E., Chiang A., Critz P., Jaklevic M.A., Lindo K., Lindquist K., Miller C., Patel S., Piacia C., Riley B.E., Rojeski H., Samiento R., Yu C., Montenegro M., Aerts A., Chung A., Abrajano A., Baker M., Gau C., Jett J., Ko C., Beall K., Woolley J.P., Stultz J.L., Kimmerly W., Martin C.H.;
RA "Sequencing of human chromosome 5."; Submitted (Feb-1998) to the EMBL/GenBank/DBJ databases.
RL -!- FUNCTION: MICROTUBULE-BASED ANTEROGRADE TRANSLATOR FOR MEMBRANOUS ORGANELLES. PLUS END-DIRECTED MICROTUBULE SLIDING ACTIVITY IN VITRO.
CC CC
CC -!- SUBUNIT: HETERODIMER OF KIF3A AND KIF3B.
CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN

```

```

CC II SUBFAMILY.
CC -!- CAUTION. Ref.3 sequence differs from that shown due to erroneous
CC gene model prediction.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF041853; AAC72294.1; --
CC DR EMBL; BC045542; AAH45542.1; --
CC DR EMBL; AC004237; AAC04475.1; AUT_SEQ.
CC DR HSSP; PL7119; 3KAR.
CC DR Genew; HGNC:6319; KIF3A.
CC DR MIM; 604683; --
CC DR GO; GO:0006996; P:organelle organization and biogenesis; TAS.
CC DR InterPro; IPR001752; kinesin_motor.
CC DR Pfam; PF00225; kinesin; 1.
CC DR PRINTS; PRO0380; KINESINHEAVY.
CC DR SMART; SM00129; KISC; 1.
CC DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
CC DR PROSITE; PS0067; KINESIN_MOTOR_DOMAIN2; 1.
CC KW Motor protein; Microtubule; ATP-binding; Coiled coil; Neurone.
CC FT DOMAIN 1 350
CC FT DOMAIN 351 599
CC FT DOMAIN 600 702
CC FT NP BIND 100 107
CC FT DOMAIN 443 446
CC FT POLY-GLU
CC FT CONFLICT 151 151
CC FT CONFLICT 170 170
CC FT CONFLICT 172 172
CC SQ SEQUENCE 702 AA; 80385 MW; 8BE5D21209B4E14 CRC64;

Query Match 38.1%; Score 664; DB 1; Length 702;
Best Local Similarity 45.8%; Pred. No. 4.5e-43;
Matches 157; Conservative 52; Mismatches 118; Indels 16; Gaps 8;

QY 6 VACVVRPLNSREESL--GETAQVYWKTDNNVYQVDGS---KSFNDRVFGHNETTK 59
D 15 VKVVVRCRPLNEREKSMCVQAVSDVMRGITTVHKTDSSNEPPKTFDFVFGESKQL 74
QY 60 NYVEIAAIPIDSALQSYNGIFAYGQTASQTYVMGSE---DHLGVIPRAIHDFOKI 116
D 75 DVYNLTARPIIDSVLLEGYNGTIFAYGQTGTGTFTMEGVRAIPELRGIPNSFAHIFGHI 134
QY 117 KKFP--DREFLRVSYMEIYNTITDLLCGTOKMKPLIREDVNRNVYVADLTVEEVYTS 175
D 135 AKAGDTRFLRVSYLEIYNEVRDLL--GKQDQORLEVKRDPDGVYIKDLSAVVNNAD 193
QY 176 MALKWITKGEKSRHYGETQNRSSRSHTIFRMILESREKGEPSNCEGSKVSHLNVDL 235
D 194 DMDRIMTLGHKRSVGATNMNHSRSHAFTITIECSEKIDGNMH--VEMGKLHLVDL 251
QY 236 AGSERAQOTGAAGVRLKEGNCINRSFILGQVKKLSDGVGFNVRDCKLTRLQNSL 295
D 252 AGSERQAKTGATGQRLKEATKINILSLTLGNVISALVDGK--STHVPYRNSKLTRLQD 310
QY 296 GGNPKTRIICTITPV--SFDETLTALQFASATKYMKNTPYV 336
D 311 GGNKTMCMCANIGPADNYDETISTLRYANRAKNIKARINE 353

RESULT 7
KF3A MOUSE
ID KF3A MOUSE STANDARD; PRT; 701 AA.
AC P28741;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Kinesin-like protein KIF3A (Microtubule plus end-directed kinesin

```

296 GGNPKTRIIICITPPV--SFDETLALQFASTAKYMNTPYNE 336
|||||::|||::|||::|||:
311 GGSKTMWCANIGPADYNDETISTLRYANRAKNIKARINE 353

RESULT 8
KF3B_HUMAN STANDARD; PRT; 747 AA.
O15066;
15-DEC-1998 (Rel. 37, Created)
15-DEC-1998 (Rel. 37, Last sequence update)
10-OCT-2003 (Rel. 42, Last annotation update)
Kinesin-like protein KIF3B (Microtubule plus end-directed kinesin motor 3B) (HH0048).
KIF3B OR KIAA0359.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RRC TISSUE=Brain;
MEDLINE=97343984; PubMed=9205841;
Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N., Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.; "Prediction of the coding sequences of unidentified human genes. VII. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro."; DNA Res. 4:141-150(1997).
[2]
SEQUENCE FROM N.A.
RP MEDLINE=21638749; PubMed=11780052;
RX Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R., Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L., Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M., Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J., Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P., Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.V., Clee C.M., Clegg S., Cobley V.E., Collier R.B., Connor R.E., Corby N.R., Coulson A., Coville G.J., Deadman R., Dhali P.D., Dunn M., Ellington A.G., Frankland J.A., Fraser A., French L., Garner P., Graham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E., Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J., Kay M.P., Kimberley A.W., King A., Knights A., Laird G.K., Lawlor S., Leharvaslatho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D., Marsh V.L., Martin S.L., McConnachie I.J., McLay K., McMurray A.A., Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T., Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I., Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H., Rice C.W., Ross M.T., Scott C.E., Sehra H.K., Showkeen R., Sims S., Skuce C.D., Smith M.H., Soderlund C., Steward C.A., Sulston J.E., Swann R.M., Symamore N., Taylor R., Tee L., Thomas D.W., Thorpe A., Tracey A., Tromans A.C., Vaubin M., Wall M., Wallis J.M., Williams S.A., Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A., Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S., Rogers J.; "The DNA sequence and comparative analysis of human chromosome 20."; Nature 414:865-871(2001).
[3]
IDENTIFICATION IN A COMPLEX WITH SMC3 AND KIFAP3B.
RP MEDLINE=98175913; PubMed=9506951;
RX Shimizu K., Shirataki H., Honda T., Minami S., Takai Y.; "Complex formation of SNAP/KAP3, a KIF3A/B ATPase motor-associated protein, with a human chromosome-associated polypeptide."; J. Biol. Chem. 273:6591-6594(1998).
RT -/- FUNCTION: Involved in tethering the chromosomes to the spindle pole and in chromosome movement. Microtubule-based anterograde translocator for membranous organelles. Plus end-directed microtubule sliding activity in vitro (By similarity).
CC -/- SUBUNIT: Heterodimer of KIF3A and KIF3B (By similarity). Interacts CC with the SMC3 subunit of the cohesin complex.


```

CC CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
CC CC II SUBFAMILY.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; AB002357; BAA20815.1; -
CC CC EMBL; AL121897; CAC16425.1; -
CC CC HSSP; P17119; 3KAR.
CC CC Genew; HGNC:6320; KIF3B.
CC CC MIM; 603754; -
CC CC GO; GO:0005873; C:plus-end kinesin complex; TAS.
CC CC GO; GO:0003777; F:microtubule motor activity; TAS.
CC CC GO; GO:0008574; F:plus-end-directed kinesin ATPase activity; TAS.
CC CC GO; GO:0008089; P:anterograde axon cargo transport; TAS.
CC CC GO; GO:0007368; P:determination of left/right asymmetry; TAS.
CC CC InterPro; IPR001752; Kinesin_motor.
CC CC Pfam; PF00225; Kinesin_1.
CC CC PRINTS; PR00380; Kinesin.
CC CC SMART; SM00129; KISC; 1.
CC CC PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
CC CC PROSITE; PS00667; KINESIN MOTOR DOMAIN2; 1.
CC CC Motor protein; Microtubule; ATP-binding; Coiled coil; Neurone.
CC CC DOMAIN 1 345 KINESIN-MOTOR (BY SIMILARITY).
CC CC FT DOMAIN 346 579 COILED COIL (BY SIMILARITY).
CC CC FT DOMAIN 580 747 GLOBULAR.
CC CC FT NP_BIND 96 103 ATP (POTENTIAL).
CC CC FT DOMAIN 386 393 POLY-GLY.
CC CC FT DOMAIN 394 406 POLY-GLU.
CC CC FT DOMAIN 723 730 POLY-SER.
CC CC SQ SEQUENCE 747 AA; 85125 MW; 97FA573AFA87023 CRC64;

Query Match 37.8%; Score 660; DB 1; Length 747;
Best Local Similarity 44.2%; Pred. No. 9.9e-43;
Matches 153; Conservative 55; Mismatches 118; Indels 20; Gaps 7;

QY 5 AVAVCVVRPLNSRESLG-----ETAQVYKTDNNVIYQVDSKSFNDFVRPHGN 55
Db 9 SVRVVVRCPMNGEKAASVDKVDVVDVKGQSVKPKGTSHM--PKTFFDAVDVWN 66

QY 56 ETTKNVVEEIAPIIDSAIQNGNTIFAYGQTASGKTYTM---MGSEDLHGVIPRAIHD 112
Db 67 AKOFELYDETFRPLVDVSLQFNGTIFAYGQTGTGKTYTMEGIRGDPKRGVIPSFDHI 126

QY 113 FQIKKFPDRFLLRVSVMEIYNETITDLCGTQKMKPLIREDVNRNRYVADLTVEVY 172
Db 127 FTHISRQNOQYLVRAVSYLEIYQBEIRDLSSKQD--TKRELKERPDTCGVYKDLSSFTVK 185

QY 173 TSEMALKWITKGEKSRHYGETKMKQSRSHITFRMILESRKGEPSKCVSVKSHLN 232
Db 186 SVKEIEHVMNVGNQNSRGATNNNEHSRSHAFVITIECEVSG--LDGENHVRGKLN 243

QY 233 VDLAGSRAAQTAGAAGVRLKGCNINSLFTILGQVVKISDQGVGFNYEDSKLRILO 292
Db 244 VDLAGSERQAKTAGGELKEATKINLSLSALGNVISALVDGK--STHLPYEDSKLRILO 302

QY 293 NSLGGNPKTRICTITVPSP--DETLALQFASHTAKMKNTPPYNE 336
Db 303 DSLGGNKTVMNVANVPASYNVEETLTTLTRVYANRAKINKPRVNE 348

RESULT 9
KF3B MOUSE
ID KF3B MOUSE STANDARD; PRT; 747 AA.
AC Q61771;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)

```

```

DE DE Kinesin-like protein KIF3B (Microtubule plus end-directed kinesin
DE DE motor 3B).
GN GN KIF3B.
OS OS Mus musculus (Mouse).
OC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX OX NCBI_TaxID=10090;
RN RN [1]
RP RP SEQUENCE FROM N.A.
RC RC STRAIN=ICR; TISSUE=Brain;
RX RX MEDLINE=96032268; PubMed=7559760;
RA RA Yamazaki H., Nakata T., Okada Y., Hirokawa N.;
RT RT "KIF3A/B: a heterodimeric kinesin superfamily protein that works as a
RT RT microtubule plus end-directed motor for membrane organelle
RT RT transport.";
RL RL J. Cell Biol. 130:1387-1399(1995).
CC CC -!- FUNCTION: Involved in tethering the chromosomes to the spindle
CC CC pole and in chromosome movement. Microtubule-based anterograde
CC CC translocator for membranous organelles. Plus end-directed
CC CC microtubule sliding activity in vitro (By similarity).
CC CC -!- SUBUNIT: Interacts with the SMC3 subunit of the cohesin
CC CC complex (By similarity). Heterodimer of KIF3A and KIF3B.
CC CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
CC CC II SUBFAMILY.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; D26077; BAA05070.1; -
CC CC PIR; A57107; A57107.
CC CC HSSP; P17119; 3KAR.
CC CC MGI; MGI:107688; Kif3b.
CC CC InterPro; IPR001752; kinesin_motor.
CC CC Pfam; PF00225; kinesin; 1.
CC CC PRINTS; PR00380; KINESINHEAVY.
CC CC SMART; SM00129; KISC; 1.
CC CC PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
CC CC PROSITE; PS00667; KINESIN MOTOR DOMAIN2; 1.
CC CC Motor protein; Microtubule; ATP-binding; Coiled coil; Neurone.
CC CC DOMAIN 1 345 KINESIN-MOTOR (BY SIMILARITY).
CC CC FT DOMAIN 346 579 COILED COIL (BY SIMILARITY).
CC CC FT DOMAIN 580 747 GLOBULAR.
CC CC FT NP_BIND 96 103 ATP (POTENTIAL).
CC CC FT DOMAIN 386 393 POLY-GLY.
CC CC FT DOMAIN 394 405 POLY-GLU.
CC CC FT DOMAIN 723 730 POLY-SER.
CC CC SQ SEQUENCE 747 AA; 85288 MW; FA369AA130EC8B47 CRC64;

Query Match 37.8%; Score 659; DB 1; Length 747;
Best Local Similarity 44.2%; Pred. No. 1.2e-42;
Matches 153; Conservative 55; Mismatches 118; Indels 20; Gaps 7;

QY 5 AVAVCVVRPLNSRESLG-----ETAQVYKTDNNVIYQVDSKSFNDFVRPHGN 55
Db 9 SVRVVVRCPMNGEKAASVDKVDVVDVKGQSVKPKGTSHM--PKTFFDAVDVWN 66

QY 56 ETTKNVVEEIAPIIDSAIQNGNTIFAYGQTASGKTYTM---MGSEDLHGVIPRAIHD 112
Db 67 AKOFELYDETFRPLVDVSLQFNGTIFAYGQTGTGKTYTMEGVRGDPKRGVIPSFDHI 126

QY 113 FQIKKFPDRFLLRVSVMEIYNETITDLCGTQKMKPLIREDVNRNRYVADLTVEVY 172
Db 127 FTHISRQNOQYLVRAVSYLEIYQBEIRDLSSKQD--TKRELKERPDTCGVYKDLSSFTVK 185

QY 173 TSEMALKWITKGEKSRHYGETKMKQSRSHITFRMILESRKGEPSKCVSVKSHLN 232
Db 186 SVKEIEHVMNVGNQNSRGATNNNEHSRSHAFVITIECEVSG--LDGENHVRGKLN 243

```

QY 233 VDLASERAAQTGAAGVRLKEGNCINRSIFILGOVKKLSGQGVGFNRYRDSKLTIRLQ 292
 Db 244 VDLASERAAQTGAAGVRLKEGNCINRSIFILGOVKKLSGQGVGFNRYRDSKLTIRLQ 302

QY 293 NSLGGNPKTRITICTIPVSF--DETALQFQASTAKYMKNTPYVNE 336
 Db 303 DSLGGNAKTVWNVGPNVVEETLTJLRANRANKNKPRVNE 348

RESULT 10
 KF4A XENLA
 ID KF4A XENLA STANDARD; PRT; 1226 AA.
 AC Q91784; Q9PS10;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Chromosome-associated kinesin KLP1 (Chromokinesin).
 GN KLP1.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, AND TISSUE
 RP SPECIFICITY.
 RC TISSUE=Oocyte;
 RX MEDLINE=95236444; PubMed=7720067;
 RA Vernos I., Raats J., Hirano T., Heasman J., Karsenti E., Wylie C.;
 RT "Xklp1, a chromosomal Xenopus kinesin-like protein essential for
 RT spindle organization and chromosome positioning.";
 RN Cell 81:117-127 (1995).
 RL [2]
 RP SEQUENCE OF 9-338 FROM N.A.
 RX MEDLINE=93246065; PubMed=8482413;
 RA Vernos I., Heasman J., Wylie C.;
 RT "Multiple kinesin-like transcripts in Xenopus oocytes.";
 RN Dev. Biol. 157:232-239 (1993).
 RL -!- FUNCTION: REQUIRED FOR MYOTIC CHROMOSOMAL POSITIONING AND BIPOLAR
 CC SPINDLE STABILIZATION.
 CC -!- SUBCELLULAR LOCATION: Nuclear. Associated with mitotic
 CC chromosomes.
 CC -!- TISSUE SPECIFICITY: Expressed in oocytes, eggs, testes and
 CC brain.
 CC -!- SIMILARITY: Belongs to the kinesin-like protein family.
 CC Chromokinesin subfamily.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X82012; CAA57539.1; --
 CC F01; I51617; I51617.
 CC HSP; P17119; 3KAR.
 CC InterPro; IPR001752; kinesin_motor.
 CC Pfam; PF00225; kinesin; 1.
 CC PRINTS; PR00380; KINESINHEAVY.
 CC SMART; SM00129; KIFC; 1.
 CC PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
 CC PROSITE; PS00667; KINESIN MOTOR DOMAIN2; 1.
 CC Motor protein; Microtubule; ATP-binding; DNA-binding;
 CC Nuclear protein; Coiled coil.
 CC FT DOMAIN 1 350 KINESIN-MOTOR.
 CC FT DOMAIN 351 1006 COILED COIL (BY SIMILARITY).
 CC FT DOMAIN 1007 1226 GLOBULAR.
 CC FT NP_BIND 87 94
 CC FT CONFLICT 163 163 I -> L (IN REF. 2).
 CC SEQUENCE 1226 AA; 138923 MW; 7F0275FCF3316697 CRC64;

Query Match 37.5%; Score 693.5; DB 1; Length 1226;
 Best Local Similarity 44.4%; Pred. No. 6.1e-42;
 Matches 154; Conservative 54; Mismatches 120; Indels 19; Gaps 9;

QY 2 BEG-AVAVCVRVPLNRSREISLG-ETAQVYWKTDNNVIYQVDSKSFNDRVHFHNETTK 59
 Db 4 DEGIPRVVALRCPLVPKNNEGCKMCLTFVPEQOVI--VGTEKSFYDYVDFPSAEQE 61
 QY 60 NYVEEIAAIPIDSAIQYNGTIFAYQTAGSKTYITMGSSEDH-----LGVIPRAIHDI 112
 Db 62 EVNSAVAPLIKGLFKGNATVLAIGYQGTSGKTYSMGGAYTHNQENEPVGVIPRVIAL 121
 QY 113 FQIKKFPDRFLLRVSYMEIYNETITDLL-CGTQKMKPLIREDNRNRYVADLTVEEV 171
 Db 122 FREIHORPEWEPFNKLVSYLEITNEEILDLLYARDKNTISIEDPKGKICGLTERDV 181
 QY 172 YTSEMALKWITKGEKSRHYGETKNQSRSSHTIFRMILESREKGPSPNCEGSKVYSHLN 231
 Db 182 KTAIDTILSCLEQNSRSTVASTAMNSQSSRSHTAIFTISIEQRKEGDKN---SFR-SKLH 237
 QY 232 LVDLASERAAQTGAAGVRLKEGNCINRSIFILGOVKKLSGQGVGFNRYRDSKLTIRI 290
 Db 238 LVDLASERAAQTGAAGVRLKEGNCINRSIFILGOVKKLSGQGVGFNRYRDSKLTIRL 297
 QY 291 LQNSLGGNPKTRITICTIPV--SFDETALQFQASTAKYMKNTPYV 335
 Db 298 LQNSLGGNPKTRITICTIPV--SFDETALQFQASTAKYMKNTPYV 344

RESULT 11
 KF17_HUMAN
 ID KF17_HUMAN STANDARD; PRT; 1029 AA.
 AC Q9P2E2; Q95077; Q8N411;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Kinesin-like protein KIF17 (KIF3-related motor protein).
 GN KIF17 OR KIF3X OR KIAA1405.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX Kimberley A., White S.;
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 78-1029 FROM N.A. (ISOFORM 1).
 RC TISSUE=Brain;
 RA Chara O., Nagase T., Kikuno R.;
 RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 239-1029 FROM N.A. (ISOFORM 1).
 RC TISSUE=Brain;
 RX MEDLINE=20181126; PubMed=10718198;
 RA Nagase T., Kikuno R., Ishikawa K.-I., Hirose M., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XVI.
 RT The complete sequences of 150 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RL DNA Res. 7:65-73 (2000).
 RN [4]
 RP SEQUENCE OF 618-1029 FROM N.A. (ISOFORM 2).
 RC TISSUE=Pancras;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,


```
FT DOMAIN 1 340 KINESIN-MOTOR (BY SIMILARITY).
FT DOMAIN 341 619 COILED COIL (BY SIMILARITY).
FT DOMAIN 620 699 GLOBULAR (BY SIMILARITY).
FT NP BIND 97 104 ATP (POTENTIAL).
SQ SEQUENCE 699 AA; 78697 MW; 7B3866111CB08190 CRC64;

Query Match 36.9%; Score 643.5; DB 1; Length 699;
Best Local Similarity 46.0%; Pred. No. 1-6e-41;
Matches 160; Conservative 49; Mismatches 114; Indels 25; Gaps 10;

QY 6 VAVCVVRPLNSREESLG-----ETAQVYWKTDNNVIYQVDGSKSFNDFRVPHG 54
DB 11 VRVVCRLPLNSKETCGFKSVVWMDMRGTGVV---TNPNA-PSGEPKPSFTFDVFP 66
QY 55 NETKNVYEIAPIIDSAIOGNGTIFAYGQTASGKTYTMMG--SEDHL-GVPIRAIHD 111
DB 67 GARQTDVYQNTARPIVDIAIEGYNGTIFAYGQTGKTFTMEGVRQPELGRGIPNSFAH 126
QY 112 IFQIKKFPDR-EFLLRVSMYIYNETITDLCCTQTKMKPLIIRVDNRNVYVADLTEEV 170
DB 127 IFGHIAKEQENVZFLRVSVLEIYNEVKDLL-GKQOHRLEVKERPVDGVVVKDLSAFV 185
QY 171 VYTSWALKWITKGEKSRHYGETKMNORSRSHITFRMILESEKGEPSNCEGSKVSHL 230
DB 186 VNNADMDRIMTLGNKRSVCGATMNNESRSHAITITLERSDMG--LDKEQHRVVGKL 243
QY 231 NVVDLAGSRAAQTGAAGVRLKGCNINRSIFILGQVVKLSGQVGGFNYRDSKLTTRI 290
DB 244 HMVDLAGSERQTKTGATGQRLKEATKINLSLSTGNVISLVDGK-STHPIYRNSKLTREL 302
QY 291 LQSLGNGNPKTRIICITTPV--SPDETTLALQFASAKYMKNTPYVNE 336
DB 303 LQSLGNGNAKIVMCANIGPAEYNYDETILSTLYANRAKNIKAKINE 350

RESULT 13
KF17 MOUSE STANDARD; PRT; 1038 AA.
ID -KF17 MOUSE
AC Q99FW8;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Kinesin-like protein KIF17 (MmKIF17).
GN KIF17.
OS Mus musculus (Mouse).
OC Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20307907; PubMed=10846156;
RA Setou M., Nakagawa T., Seog D.-H., Hirokawa N.;
RT "Kinesin superfamily motor protein KIF17 and mLin-10 in NM2A
RT receptor-containing vesicle transport.";
RL Science 288:1796-1802(2000).
CC -!- FUNCTION: Transports vesicles containing N-methyl-D-aspartate
CC (NMDA) receptor 2B along microtubules.
CC -!- SUBUNIT: Interacts with LIN-10 PDZ domain.
CC -!- TISSUE SPECIFICITY: Neuronal-specific.
CC -!- SIMILARITY: Belongs to the kinesin-like protein family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB008867; BAB21099.1; -.
CC DR HSSP; P17119; 3KAR.
CC MGD; MGI:1098229; Kif17.
DR
```

```
DR GO; GO:0005871; C:kinesin complex; IDA.
DR GO; GO:0003777; F:microtubule motor activity; IDA.
DR GO; GO:0005515; F:protein binding; IPI.
DR GO; GO:0007017; P:microtubule-based process; IDA.
DR GO; GO:0016192; P:vesicle-mediated transport; IDA.
DR InterPro; IPR001752; kinesin_motor.
DR Pfam; PF00225; kinesin; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
DR PROSITE; PS00667; KINESIN MOTOR DOMAIN2; 1.
KW Motor protein; Microtubule; ATP-binding; Coiled coil; Transport;
KW Protein transport. 265 KINESIN-MOTOR.
FT DOMAIN 1 265 COILED COIL (POTENTIAL).
FT DOMAIN 346 470 COILED COIL (POTENTIAL).
FT DOMAIN 748 855 ATP (POTENTIAL).
FT NP BIND 91 98 ATP (POTENTIAL).
SQ SEQUENCE 1038 AA; 116372 MW; 2BED852A3AFBD46 CRC64;

Query Match 36.8%; Score 641.5; DB 1; Length 1038;
Best Local Similarity 42.5%; Pred. No. 4e-41;
Matches 147; Conservative 59; Mismatches 119; Indels 21; Gaps 8;

QY 5 AVAVCVVRPLNSREESLG-ETAQVYWKTDNNVIYQVDG-----KSFNDFRVPHGNETT 58
DB 5 SVKVVVRCRPMNKRERELSCQSVVTVDSARGQCFCIQNPGAADPPKQFTFDGAYIEHFT 64
QY 59 KNYVEIAPIIDSAIOGNGTIFAYGQTASGKTYTMMGSED--HLGVPIRAIHDFOK 115
DB 65 EQIYNEIAYPLVGVTEGYNGTIFAYGQTGSGKSFMTQGLPDPFCQGIIPRAFEHVFS 124
QY 116 IKFPPDEFLRVSMYIYNETITDLCCTQTKMKPLIIRVDNRNVYVADLTEEVVTS 175
DB 125 VQCAENTKFLVRASYLEIYNEVDHLL-GADTKQRLKELKEHPEKGVVYVKGSLSMHTVHVA 183
QY 176 MALKWITKGEKSRHYGETKMNORSRSHITFRMILE---SREKGEPSNCEGSKVSHLNL 232
DB 184 QCERVMTGKNRAVGTYTLMKDSRSRSHSIFTINIEIYAVDERG-----KDHLAGKLN 238
QY 233 VDLAGSRAAQTGAAGVRLKGCNINRSIFILGQVVKLSGQVGGFNYRDSKLTTRI 292
DB 239 VDLAGSERQTKTGATGQRLKEATKINLSLSTGNVISLVDGRC-KHPIYRDSKLTREL 297
QY 293 NLSGNGNPKTRIICITTPV--SFDETTLALQFASAKYMKNTPYVNE 336
DB 298 DSLGNGYKTLMWACLSPADNNYDETILSTLYANRAKNIKAKINEPRINE 343

RESULT 14
OSM3_CABEL STANDARD; PRT; 672 AA.
ID -OSM3_CABEL
AC P46873;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Kinesin-like protein osm-3.
GN OSM-3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodexinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=95230679; PubMed=7714894;
RA Tabish M., Siddiqui Z.K., Nishikawa K., Siddiqui S.S.;
RT "Exclusive expression of C. elegans osm-3 kinesin gene in
RT chemosensory neurons open to the external environment.";
RL J. Mol. Biol. 247:377-389(1995).
RN [2]
RP PRELIMINARY SEQUENCE OF 1-397 FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=93379214; PubMed=7690265;
DR
```

RA Shalir M.A., Fukushige T., Yasuda H., Miwa J., Siddiqui S.S.;
 RT "C. elegans oem-3 gene mediating osmotic avoidance behaviour encodes
 RL a kinesin-like protein.";
 CC NeuroReport 4:891-894(1993).
 CC -!- TISSUE SPECIFICITY: AMPHID AND IL2 NEURONS.
 CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
 CC II SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D38632; BAA07612.1; -;
 DR EMBL; D14968; BAA20996.1; -;
 DR PIR; S54351; S54351.
 DR HSP; P17119; 3KAR.
 DR InterPro; IPR001752; kinesin_motor.
 DR PRINTS; PR00380; KINESINHEAVY.
 DR SMART; SM00129; KISC; 1.
 DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
 DR PROSITE; PS00667; KINESIN MOTOR DOMAIN2; 1.
 KW Motor protein; Microtubule; ATP-binding; Coiled coil.
 FT DOMAIN 57 300 KINESIN-MOTOR.
 FT NP_BIND 461 497 COILED COIL (POTENTIAL).
 FT NP_BIND 130 137 ATP (POTENTIAL).
 SQ SEQUENCE 672 AA; 75456 MW; 4E2160F7042AFCD7 CRC64;
 Query Match 36.3%; Score 633; DB 1; Length 672;
 Best Local Similarity 47.0%; Pred. No. 9, 9e-41;
 Matches 142; Conservative 49; Mismatches 97; Indels 14; Gaps 6;
 QY 41 DG-SKSNFDRVFNGETTKNVYBEIAPIIDSIAIQYNGTIFAYGQTASGKTYTMMGE 99
 DB 85 DGAADFTDFGAYFMIRPGEQIYNDIVFPLVENIEGYNGTIFAYGQTGSKTSMQIE 144
 QY 100 D---HLGVPIRAIHDFQIKKFPDRELLRVSYMEIYNEITDILCGTQKMKLIRED 156
 DB 145 TIPQRGVIPAFPHITATATTENVPFLVHCSEIYNEVRLDGLADNKKQ-LEIKQ 203
 QY 157 VNRNVYVADLVEEYVYVSEMALKWITKGEKSRHYGETKNORSSRSHITFPMILESREK 216
 DB 204 PDGRGVYVAGLSMHVCHDVPACKELMTRGFNNRHVGATLMKDSRSRSHIFVYVEGITE- 262
 QY 217 EPSNCEGSVKVSHNLVDLAGSRAAOTGAAGVRLKEGCNINRSLFTLGQVYKLSGQV 276
 DB 263 -----TGSIRMGKLNLDLAGSERQSKTGATGDRLEATKINLSALGNVISALVDGK- 316
 QY 277 GGFYNYRDSKLTILQNSLGNPKTRILCTIPYS--FDETLTALQASTAKWNTPYV 334
 DB 317 SKHIPYRDSKLTLLQSLGNTKTIIMACVSPSSDNYDETLSTLRVANKNIKNKPTI 376
 QY 335 NE 336
 DB 377 NE 378
 RESULT 15
 KF4A CHICK STANDARD; PRT; 1225 AA.
 AC Q90640; Q90608;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Chromosome-associated kinesin KIF4A (Chromokinesin).
 GN KIF4A.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, AND TISSUE
 RP SPECIFICITY.
 RC STRAIN=White leghorn; TISSUE=Embryonic retina;
 RX MEDLINE=95181533; PubMed=7876303;
 RA Wang S.Z., Adler R.;
 RT "Chromokinesin: a DNA-binding, kinesin-like nuclear protein.";
 RL J. Cell Biol. 128:761-768(1995).
 RN [2]
 RP SEQUENCE OF 728-1088 FROM N.A.
 RC STRAIN=White leghorn; TISSUE=Embryonic retina;
 RX MEDLINE=94151228; PubMed=8108415;
 RA Wang S.Z., Adler R.;
 RT "A developmentally regulated basic-leucine zipper-like gene and its
 RT expression in embryonic retina and lens.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:1351-1355(1994).
 CC -!- FUNCTION: REQUIRED FOR MITOTIC CHROMOSOMAL POSITIONING AND BIPOLAR
 CC SPINDLE STABILIZATION.
 CC -!- SUBCELLULAR LOCATION: Nuclear. Associated with mitotic
 CC chromosomes.
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN PROLIFERATING CELLS;
 CC NEUROEPITHELIUM OF EMBRYOS.
 CC -!- SIMILARITY: Belongs to the kinesin-like protein family.
 CC Chromokinesin subfamily.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U18309; AAC59666.1; -;
 DR EMBL; U04821; AAA18960.1; -;
 DR PIR; A56514; A56514.
 DR HSP; P17119; 3KAR.
 DR InterPro; IPR001752; kinesin_motor.
 DR PRINTS; PR00225; kinesin; 1.
 DR PRINTS; PR00380; KINESINHEAVY.
 DR SMART; SM00129; KISC; 1.
 DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
 DR PROSITE; PS00667; KINESIN MOTOR DOMAIN2; 1.
 KW Motor protein; Microtubule; ATP-binding; DNA-binding;
 KW Nuclear protein; Coiled coil.
 FT DOMAIN 1 351 KINESIN-MOTOR.
 FT DOMAIN 352 1003 COILED COIL (BY SIMILARITY).
 FT DOMAIN 1004 1225 GLOBULAR.
 FT NP_BIND 88 95 ATP (POTENTIAL).
 FT CONFLICT 1087 1088 KG -> RI (IN REF. 2).
 SQ SEQUENCE 1225 AA; 138923 MW; FA01ED83425F5875 CRC64;
 Query Match 36.3%; Score 633; DB 1; Length 1225;
 Best Local Similarity 42.8%; Pred. No. 2, 2e-40;
 Matches 148; Conservative 59; Mismatches 121; Indels 18; Gaps 8;
 QY 2 EGAVAVCVVRPLNSREESLG-ETAQVYWKTDNNVIYQVDSKSNFDRVFNHNETTKN 60
 DB 6 EKGIPVRVVRCPPLVPKETSEGQCMCLSFVPGPQVI--VGSKATYDVDFPSVEQEE 63
 QY 61 VYEIAAPIDSIAIQYNGTIFAYGQTASGKTYTMMG----SDH---LGVIPRAIHDF 113
 DB 64 VENTAVAPLIRGIFKGNATVLAAYGQTGSKTYSMGTYTASQEHDPDMGVIRVVKLLF 123
 QY 114 QKIKFPDRELLRVSYMEIYNEITDILCGT-QKMKPLIREDVNRVNTVADLTVEVY 172
 DB 124 KEKEQDWEFVLKVSYLEIYNEITDILDLCSRSRSQISIREDPKEGKIVGLTERNA 183
 QY 173 TSEMALKWITKGEKSRHYGETKNORSSRSHITFPMILESREKGEPSNCEGSVKVSHNL 232
 DB 184 SARDTVSCLEQGNCHTAVSTANSSSRSHAITFICIDQKKK----NDKNSFHSKJHL 239

Fri Aug 6 10:49:09 2004

Qy	233	VDLAGSERAAQTGAAGVRLKEGCNINRSIFILGOVKKI-SDGOVGGFINYRDSKLTRIL	291
Db	240	VDLAGSERQKKTAEGRDLKEGININRGLLCLGNVISALGEENKGGFVYRDSKLTRIL	299
Qy	292	QNSLGGNPKTRIICTITPV--SFDETLTALQFASTAKYMKNTPYVN	335
Db	300	QDSLGGNSHTLMIACVSPADSNLEETLNTIRYADRARKIKKPIVN	345

Search completed: July 29, 2004, 09:36:14
Job time : 4.25614 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 09:27:31 ; Search time 14.3994 Seconds

(without alignments)
7428.155 Million cell updates/sec

Title: US-10-045-631B-88_COPY_2_340

Perfect score: 1744

Sequence: 1 AEEGAVAVCVVRPLNSREE.....QFASTAKYMKNPYYNEVST 339

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25.*

1: sp_archaea.*

2: sp_bacteria.*

3: sp_fungi.*

4: sp_human.*

5: sp_invertebrate.*

6: sp_mammal.*

7: sp_mhc.*

8: sp_organelle.*

9: sp_phage.*

10: sp_plant.*

11: sp_rodent.*

12: sp_virus.*

13: sp_vertebrate.*

14: sp_unclassified.*

15: sp_rvirus.*

16: sp_bacteriap.*

17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1538	88.2	549	11 Q7TPX4	Q7tpx4 mus musculus
2	1299.5	74.5	2954	13 Q42263	Q42263 xenopus lae
3	795.5	45.6	807	10 Q94HV9	Q94hv9 arabidopsis
4	795.5	45.6	823	10 Q9S7P3	Q9s7p3 arabidopsis
5	767.5	44.0	459	10 Q9S830	Q9s830 arabidopsis
6	760	43.6	888	10 Q9LQ62	Q9lq62 arabidopsis
7	734	42.1	1885	5 Q869B8	Q869b8 dictyosceli
8	707	40.5	160	11 Q35059	Q35059 mus musculus
9	707	40.5	2013	5 Q9VK10	Q9vki0 drosophila
10	707	40.5	2244	5 Q9NC60	Q9nc60 drosophila
11	686	39.3	1055	10 Q8RW4	Q8rw4 arabidopsis
12	686	39.3	1055	10 Q8W5R5	Q8w5r5 arabidopsis
13	680	39.0	699	5 Q8MPT8	Q8mpt8 caenorhabdi
14	674	38.6	1058	10 Q9SJT0	Q9sjt0 arabidopsis
15	669.5	38.4	1033	10 Q9LHL9	Q9lhl9 arabidopsis
16	668	38.3	671	4 Q86XX7	Q86xx7 homo sapien

17	668	38.3	1127	4 Q86TN3	Q86tn3 homo sapien
18	667.5	38.3	954	10 Q9AWM8	Q9awm8 oryza sativ
19	666.5	38.2	1459	3 P87198	P87198 ustilago ma
20	665.5	38.2	819	11 Q7TQG6	Q7tqg6 mus musculu
21	665.5	38.2	1231	11 Q80YP3	Q80yp3 mus musculu
22	664	38.1	408	4 Q81WH8	Q81wh8 homo sapien
23	663	38.0	408	11 Q80UK1	Q80uk1 mus musculu
24	663	38.0	443	11 Q8CGJ1	Q8cgj1 mus musculu
25	663	38.0	701	11 Q7TSZ7	Q7tsz7 mus musculu
26	663	38.0	959	10 Q8S950	Q8s950 nicotiana t
27	659	37.8	699	13 Q98T11	Q98t11 xenopus lae
28	659	37.8	747	11 Q8BNH4	Q8bnh4 mus musculu
29	659	37.8	757	11 Q80U27	Q80u27 mus musculu
30	657	37.7	890	10 Q8W5R6	Q8w5r6 arabidopsis
31	655.5	37.6	836	5 Q7YUC7	Q7yuc7 tetrahymena
32	650.5	37.3	735	5 Q9U0D5	Q9u0d5 tetrahymena
33	648.5	37.2	956	10 Q9C7B9	Q9c7b9 arabidopsis
34	648.5	37.2	974	10 Q8S905	Q8s905 arabidopsis
35	646.5	37.1	677	5 Q9VRK9	Q9vrk9 drosophila
36	645	37.0	671	5 Q8MPT7	Q8mpt7 caenorhabdi
37	645	37.0	672	5 Q9GV93	Q9gv93 caenorhabdi
38	643.5	36.9	997	10 Q9FG03	Q9fg03 arabidopsis
39	638.5	36.6	782	5 Q19633	Q19633 caenorhabdi
40	638.5	36.6	782	5 Q9GV92	Q9gv92 caenorhabdi
41	633.5	36.3	695	13 Q802X4	Q802x4 brachydanio
42	633	36.3	932	10 Q9LXL3	Q9lxl3 arabidopsis
43	633	36.3	937	10 Q8LGU3	Q8lgu3 arabidopsis
44	633	36.3	937	10 Q8LSJ2	Q8lsj2 arabidopsis
45	633	36.3	938	10 Q8LNZ2	Q8lnz2 arabidopsis

ALIGNMENTS

RESULT 1

Q7TPX4	PRELIMINARY;	PRT;	549 AA.
AC Q7TPX4;			
DT 01-OCT-2003 (Tremblrel. 25, Created)			
DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)			
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)			
DE Hypothetical protein (Fragment).			
OS Mus musculus (Mouse).			
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX NCBI_TaxID=10090;			
RN [1]			
RC SEQUENCE FROM N.A.			
RP STRAIN=C57BL/6J; TISSUE=Egg;			
RX MEDLINE=22389257; PubMed=12477932;			
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,			
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,			
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,			
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,			
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,			
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,			
RA Brownstein M.J., Udell T.B., Toshylyuk S., Carninci P., Prange C.,			
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,			
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,			
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,			
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,			
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,			
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,			
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,			
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,			
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,			
RA Jones S.J., Marra M.A.			
RC "Generation and initial analysis of more than 15,000 full-length human			
RT and mouse cDNA sequences."			
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).			
RN [2]			
RP SEQUENCE FROM N.A.			
RC STRAIN=C57BL/6J; TISSUE=Egg;			

RA	Strausberg R.;	Submitted (May-2003) to the EMBL/GenBank/DBDJ databases.
RL	EMBL; BC052843; AAH52843.1; -	
DR	Hypothetical protein.	
KW	NON TER 549 549	
FT	SEQUENCE 549 AA; 62899 MW; B707C97DC9FB38D3 CRC64;	
SEQ		
	Query Match	88.2%; Score 1538; DB 11; Length 549;
	Best Local Similarity	88.2%; Pred. No. 3.6e-113;
	Matches 298; Conservative 11; Mismatches 29; Indels 0; Gaps 0	
Qy	1	AEEGAVAVCVVRPLNSREESIGETAQVYWKTDNNVYQVGSKSFNDRVFHGNETTKN 60
Db	2	AEERAVAVCVVRPLNSREELGEATHYWKDKNAYIQSDGKSFQDRVDFDSNETTKN 61
Qy	61	VYEETAAPIIDSAIQGYNGTTFAYQOTASGKTYTWMGSEDLHGLVTPRAIHDFQIKKPP 120
Db	62	VYEETAVPIISALQGYNGTTFAYQOTASGKTHTWGSEDLGLVTPRAIHDFQIKKPP 121
Qy	121	REFLLRVSYMEIYNETTDLILCGTKMKPLIREDVNVNVTADITREVVYTSMAWK 180
Db	122	BREFLLRVSYMEIYNETTDLICNAQMKPLIREDTNKTVVSDITREVVYTSMAWK 181
Qy	181	ITWGEKSHYGETQKNQSRSSHTTFRMILESRKEGPEPNCESGVKVSHLNVLVDLAGSR 240
Db	192	LATGEKNRHYGITKNQSRSSHTTFRMILESRKAEPSNCDGSKVVSHLNVLVDLAGSR 241
Qy	241	AAQTGAAGVRLKEGCNINRESFILGVTKLSDGVGGFINVRDYSKLTILQNSLGGNPK 300
Db	242	AAQTGAAGVRLKEGCFINNLFILGVTKLSDGVGGFINVRDYSKLTILQNSLGGNAK 301
Qy	301	TRIIITTPVSPDETLTALQFATAKMYNTPYNVEVS 338
Db	302	TRIIITTPASIDETLTALQFATAKMYNTPYNVEVS 339

```

RESULT 2
CD2263
ID O42263 PRELIMINARY; PRT; 2954 AA.
AC O42263;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Kinesin-related protein.
GN XCENP-E.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=98028574; PubMed=9363944;
RA Wood K.W., Sakowicz R., Goldstein L.S., Cleveland D.W.;
RT "CENP-E is a plus end-directed kinetochore motor required for
RT metaphase chromosome alignment.";
RL Cell 91:357-366(1997).
DR BWBL; AF027728; AAC60300.1; -.
DR PIR; T14156; T14156.
DR HSP; P33176; I8G2.
DR GO; GO:0005871; C:kinesin complex; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR GO; GO:0007017; P:microtubule-based process; IEA.
DR InterPro; IPR001752; kinesin_motor.
DR Pfam; PF00225; kinesin; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SMO0129; KISC; 1.
DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
DR PROSITE; PS50067; KINESIN MOTOR DOMAIN2; 1.
SQ SEQUENCE 2954 AA; 339964 MW; 439804ED08592679 CRC64;
Query Match 74.5%; Score 1299.5; DB 13; Length 2954;

```

Query Match


```

Db 55 TNASVVELLT KDIIHAAVEGFGNGTAFAYGQTSSGKFTTGTSETDPIIRSRVDVFERI 114
QY 117 KFPDPREFLLRVSYMEIYNETITDILCGTQKWKPLIIRDVNRNVVADLTREVVYTS 176
Db 115 HMSIDREFLLRVSYMEIYNEEINDLL--AVENQRLOIHEHLRGVFAGLKEEIVSDAQ 172
QY 177 ALKWTITKESKSHYGETKMNQSSRSHTTIFRMILSRKGEPSNCEGSKVSHLNLVDLA 236
Db 173 ILKLDGSEVNRHFGFTNMNVHSSRSHTTIFRMVIESR--GKDNSSSDAIRSVLNLVDLA 230
QY 237 GSERAQOTCAAGVRLKEGNCINRSFILGOVTKKLSDG--QVGGFNYRDSKLTIRLONSL 295
Db 231 GSERAKTGAGVRLQEGKYINKSLMILGNVINKLSDSTKLAHPTPYRDSKLTIRLOPAL 290
QY 296 GGNPKTRIICTIPVS--FDETLTALQFASTAKYMKNTPYNEVST 339
Db 291 GGNAKTCIICTIAPBEHHIESKGTLPFASRAKRITNCAQVNEILT 336

RESULT 4
Q9S7P3 PRELIMINARY; PRT; 823 AA.
AC Q9S7P3
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE 01-OCT-2003 (TREMBlrel. 25, Last annotation update).
DE Kinesin-like protein.
GN ZCF125.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Ooe H., Kato A., Kameda Y.;
RT "Arabidopsis thaliana genomic sequence for a kinesin-like protein.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Kato A., Suzuki M., Kuwahara A., Ooe H., Higano-Inaba K., Kameda Y.;
RT "Isolation and analysis of cDNA within a 300 kb Arabidopsis thaliana
RL genomic region located around the 100 map unit of chromosome 1.";
RL Gene 239:309-316(1999).
DR EMBL; AB028470; BAB88114.1; -
DR EMBL; AB028468; BAB88112.1; -
DR PIR; T52425; T52425.
DR HSSP; P33176; 1BG2.
DR GO; GO:0005871; C:kinesin complex; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR GO; GO:0007017; P:microtubule-based process; IEA.
DR InterPro; IPR001752; Kinesin_motor.
DR Pfam; PF00225; kinesin; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
DR PROSITE; PS00667; KINESIN MOTOR DOMAIN2; 1.
SQ SEQUENCE 823 AA; 93148 MW; 6AFB1C622E4632C9 CRC64;

Query Match 45.6%; Score 795.5; DB 10; Length 823;
Best Local Similarity 51.4%; Pred. No. 3.8e-54;
Matches 178; Conservative 48; Mismatches 95; Indels 25; Gaps 7;

QY 6 VAVCVVRPLNRSREESIGETAQVYWKNTDNNVYQVDSKSFN-----FDRVHGNE 56
Db 4 ICVAVRVP-----PAPENGASLWKVEDN---RISLHKSLLDTPITASHAFDHFVDESS 54
QY 57 TTKNVYEEIAAPLIDSAIQYNGTIFAYGQTASGKTYTMGSEDHGLGVIPRAHIFQKI 116

```

```

Db 55 TNASVVELLT KDIIHAAVEGFGNGTAFAYGQTSSGKFTTGTSETDPIIRSRVDVFERI 114
QY 117 KFPDPREFLLRVSYMEIYNETITDILCGTQKWKPLIIRDVNRNVVADLTREVVYTS 176
Db 115 HMSIDREFLLRVSYMEIYNEEINDLL--AVENQRLOIHEHLRGVFAGLKEEIVSDAQ 172
QY 177 ALKWTITKESKSHYGETKMNQSSRSHTTIFRMILSRKGEPSNCEGSKVSHLNLVDLA 236
Db 173 ILKLDGSEVNRHFGFTNMNVHSSRSHTTIFRMVIESR--GKDNSSSDAIRSVLNLVDLA 230
QY 237 GSERAQOTCAAGVRLKEGNCINRSFILGOVTKKLSDG--QVGGFNYRDSKLTIRLONSL 295
Db 231 GSERAKTGAGVRLQEGKYINKSLMILGNVINKLSDSTKLAHPTPYRDSKLTIRLOPAL 290
QY 296 GGNPKTRIICTIPVS--FDETLTALQFASTAKYMKNTPYNEVST 339
Db 291 GGNAKTCIICTIAPBEHHIESKGTLPFASRAKRITNCAQVNEILT 336

RESULT 5
Q9SS30 PRELIMINARY; PRT; 459 AA.
AC Q9SS30
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Putative kinesin-like centromere protein.
GN F14P13.22.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,
RA Ronning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,
RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RT "Arabidopsis thaliana chromosome III BAC F14P13 genomic sequence.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC009400; AAF02823.1; -
DR HSSP; P33176; 1BG2.
DR GO; GO:0005871; C:kinesin complex; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR GO; GO:0007017; P:microtubule-based process; IEA.
DR InterPro; IPR001752; Kinesin_motor.
DR Pfam; PF00225; kinesin; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
DR PROSITE; PS00667; KINESIN MOTOR DOMAIN2; 1.
SQ SEQUENCE 459 AA; 51872 MW; 7C8487E9B7038E6A CRC64;

Query Match 44.0%; Score 767.5; DB 10; Length 459;
Best Local Similarity 55.5%; Pred. No. 2.7e-52;
Matches 166; Conservative 45; Mismatches 77; Indels 11; Gaps 5;

QY 49 DRVFHGNETTKNVYEIAAPLIDSAIQYNGTTFAYGQTASGKTYTMGSEDHGLGVIPRA 108
Db 45 DRIFREDCKTVQVYEARTKEIVSAARGVNGTVFAIGQTNSTGKTHTWGSPFIEPGVIPLA 104
QY 109 IHDIFOKIKKPPDREFLLRVSYMEIYNETITDILCGTQKWKPLIIRDVNRNVVADLTRE 168
Db 105 VHDLFDTIYQDASREFLLRMSYLEIYNEDINDLL--APEHRKLIQHLENLEKGFVAGLRE 162
QY 169 EVVYTTSEMALKWITKESKSHYGETKMNQSSRSHTTIFRMILSRK-----GPSNCEGS 224
Db 163 EIVASPPQVLEWMEFGESEHRRHIGETNMNLYSSRSHTTIFRMILSRKQMDGEGVNSCD-A 221
QY 225 VKVSHLNLVDLAGSRAAQTGAAGVRLKEGNCINRSFILGOVTKKLSDG--QVGGFNY 292

```


Matches	166;	Conservative	51;	Mismatches	103;	Indels	16;	Gaps	8;
Qy	5	AVAVCVRVRLPLNSREESLGETAQVYWK-T-DNNVIYQVDG-SKSFNFDPRVFHGNETTKNY	62						
Ddb	8	SIQVCIVKVP-CP-EGTSL-----WQVKERSIHLADSHAEPPYVDYVDFDEGASNQEVF	60						
Qy	63	EIIAAPIIDSALIQGVNGTIIPAYGQTASGYTYTMGSEDLGLVIPRAIHDI FQIKKKFPDR	122						
Ddb	61	DRMAKHIVHACMQGGNGTIFAYGQTSSGKYTYTMGDEQPGVMVLAAKEIFOQISSETER	120						
Qy	123	EFLLRVSWEINYEITIDLLCGTQKMKPLIIDEDVRNVVVADLTDEVVYTSEM-ALKWI	181						
Ddb	121	DFLLRGYTEIEYNKYTDLL--NKQNQDLKIHESGNGVINVN--CBECIITSVDLRL	176						
Qy	182	TGGEKSRHYGETKMNRORSRSHTFIRMILESREKGPFSCBGSVKVSHMLNLDLAGSERA	241						
Ddb	177	CLGNKERTVGETWMNRSRSHAIFKLIESR-KSDHSD-DDAVIQSVLNLDLAGSERA	234						
Qy	242	AQTGAAGVRLKEGCNINRSLFILGOVKKLSDGQGVGFNRYRDSKLTRILONSLGQNPKT	301						
Ddb	235	DQTGARGARLKEGGHINKSLFLSNVTKLSENADRFTNYRDSKLTRILOASLGNGNAFT	294						
Qy	302	RICCTTPYSFDELTAQLQASTAKYMKNTPYNEV	337						
Ddb	295	SIICTKPIMEBSQSTLSLFATRAKKIRIKPQVNM	330						

RESULT 10

ID	Q9NCGO	PRELIMINARY;	PRT:	2244 AA.
AC	Q9NCGO;			
DT	01-OCT-2000 (TrEMBLrel. 15, Created)			
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)			
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)			
DE	Kinesin-like kinetochore motor protein CENP-meta.			
GN	CNET OR CG6392.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;			
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephydroidea; Drosophilidae; Drosophila.			
OX	NCHI_Taxid=7227;			
RX	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20351410; PubMed=10893249;			
RA	Yucel J.K., Marszalek J.D., McIntosh J.R., Goldstein L.S.B.,			
RA	Cleveland D.W., Philp A.V.			
RT	"CENP-meta, an Essential Kinetochore Kinesin Required for the			
RL	Maintenance of Metaphase Chromosome Alignment in Drosophila.";			
RL	J. Cell Biol. 150:1-12(2000).			
DR	EMBL; AF220353; AAF32355.1; -.			
DR	HSP; P31176; IBC2.			
DR	FlyBase; FBgn0040232; cmet.			
DR	GO; GO:0005639; C:kinetochore; IDA.			
DR	GO; GO:0007080; P:mitotic metaphase plate congression; IMP.			
DR	InterPro; IPRO01752; kinesin_motor.			
DR	Pfam; PF00225; kinesin; I.			
DR	PRINTS; PR00380; KINESINHEAVY.			
DR	SMART; SM00129; KISC; 1.			
DR	PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.			
DR	PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.			
SO	SEQUENCE 2244 AA; 257992 MW; FA6A3AB2A541ADE0 CRC64;			

Query Match	40.5%;	Score	707;	DB	5;	Length	2244;
Best Local Similarity	49.4%;	Pred.	No. 1.7e-46;				
Matches	166;	Conservative	51;	Mismatches	103;	Indels	16;
Gaps							

Qy	5	AVAVCRVRLPLNSREESIGETAAQVYWK	-DNNVTVQVDG	-SKSFNPDRVPHGNETKTVY	62
	::	::	::	::	::
	::	::	::	::	::
Db	8	SIQVCIKVRCPBPGTSL	-----WQVKEGRSIIHLSHAPFYDYVDFEGASNQEVF	60	
Qy	63	REIAAPIDLSAQVNGNTIFAYGQTASGKTYTMGSEDLHGVTPRAIHDF	FQIKKKFPDR	122	
	61	DEMREBETVHAQWGFNETIFAYGQTSNGKTYTMGDEONPGVYVLAKEKIF	FOQISSETER	120	

QY	123	PELLRVSYMEIYNETITDLLCGTQMKPLIIEEDVNRNYYVADLTVEVVYTSQM-ALKWI	181
DB	121	DELLRVGVIEIYNEKIYDLL--NKKQDLKHESGNGVNVN--CBECIITSEVDLRL	176
QY	182	TKGEXRAYGETKMNQSRSHITFRMILESEKGEPSNCEGSVKYSHNLNVDLAGSERA	241
DB	177	CLGNKERTVGETNMNERSRSHAIFKIITESR-KSDHSD-DDAVIOSVNLNVDLAGSERA	234
QY	242	AQTGAAGVPLKGCNINSLFILGOVYIKKLDQGVGFYINRYSKLTIRLQNSLGGNPKT	301
DB	235	DOTGARGALKKGGHINKSLFLNSVIKSELSEADNRFTNYRDSKLTIRLQASLGGNAFT	294
QY	302	RIICITPVSFDETLTALQFASHTAKYMNKNTPPYNEV	337
DB	295	SIICITKPSIMEESQSTLSFATRAKKIRIKPQVNEM	330
RESULT	11		
Q8RW4			
ID	Q8RW4	PRELIMINARY; PRT; 1055 AA.	
AC	Q8RW4;		
DT	01-JUN-2002 (Tremblrel. 21, Created)		
DT	01-JUN-2002 (Tremblrel. 21, Last sequence update)		
DT	01-OCT-2003 (Tremblrel. 25, Last annotation update)		
DE	Putative kinesin.		
GN	AT4G39050.		
OS	Arabidopsis thaliana (Mouse-ear cress).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;		
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsi.		
OX	NCBI_TaxID=3702;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M.,		
RA	Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,		
RA	Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,		
RA	Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,		
RA	Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,		
RA	Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,		
RA	Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Shinozaki K.,		
RA	David R.W., Ecker J.R., Theologis A.;		
RT	"Arabidopsis Full Length cDNA Clones.;"		
RL	Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.		
DR	EMBL: AY091060; AAM1381.1; -		
DR	GO: GO:0005871; C:kinesin complex; IEA.		
DR	GO: GO:0005524; P:ATP binding; IEA.		
DR	GO: GO:0003774; F:motor activity; IEA.		
DR	GO: GO:0007017; P:microtubule-based process; IEA.		
DR	InterPro: IPR001752; kinesin_motor.		
DR	InterPro: IPR001841; Znf ring.		
DR	Pfam: PF00225; kinesin; I.		
DR	PRINTS: PR00380; KINESINHEAVY.		
DR	SMART: SM00129; KISC; I.		
DR	SMART: SM00184; RING; I.		
DR	PROSITE: PS00411; KINESIN MOTOR DOMAIN1; 1.		
DR	PROSITE: PS00067; KINESIN MOTOR DOMAIN2; 1.		
DR	PROSITE: PS50089; ZF RING2; 1.		
CC	CROUZE 1955 AA..116452 MW: BBF3709ACB40215B CRC64;		

	Query Match	39.3%	Score 686;	DB 10;	Length 1055;
	Best Local Similarity	46.2%	Pred. No. 2.6e-45;		
	Matches 159;	Conservative 57;	Mismatches 114;	Indels 14;	Gaps 7
QY	1	AEEGAVAVCVVRPLNLSREESLGTAQVYWKTDNNVIVQDGSKSFNFQRFVPHGNETTKN	60		
DB	94	SERDSISVTVFRPLSDREYQRGDEVAVYPDGUTLVRHEYNLPTAYAFDKVGPQQTATID	153		
QY	61	VYEEIAAPTIIDSAIOGYNGTIFAYGQTASGTYITWMSGDHGLVPIRAIHDLFQKIKKEP	120		
DB	154	VYDVAARPVWKAAMEGVNGTVFAYGVTSSTGKTHMHGDQESPCIIPLAIKDFVSIIQDTP	213		
QY	121	DREFTLLRSVMEIYNETITDLLCGTQKQKPLIIRDVNRNVVADLTEEVVYTSSEMALKW	180		

Db 214 GREFFLRVSYLRIYNEVINDLLDPTG--QNLRVRED--SQGTVEGKKEEVLSPGHLSF 270
 QY 181 IYKGSKRHYGTQKMNORSRSHITFRMILESGKEPNCESGVKVSHLNLDLAGSER 240
 Db 271 IAAGEHRHVGNFNLLSRSHITFTLMVESATGDEYD--GVTFSQLNLDLAGSE- 326
 QY 241 AAQTGAAGVRLKEGNCNINRSLFQVKKLSDGVGGFINTYRDSKLTILQNSLGGNPK 300
 Db 327 SSKTETGLRRKEGYSINKSLTLGTVICKLSEK-ATHIPYRDSKLTLLQSSLSGHGH 385
 QY 301 TRIICTITPV--SPDETITLQFASAK-----YMKNTPTVNEVS 338
 Db 386 VSLICTITPASSSSEETHNTLKFASRAKSIETIYASRNQIIDEKS 429

RESULT 12

QW5R5 PRELIMINARY; PRT; 1055 AA.
 AC QW5R5;
 DT 01-MAR-2002 (TREMELrel. 20, Created)
 DT 01-OCT-2003 (TREMELrel. 25, Last sequence update)
 DE Kinesin-related protein (Putative kinesin protein).
 GN MKP2 OR A74G39050.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21563048; PubMed=11706156;
 RA Itoh R., Fujiwara M., Yoshida S.;
 RT "Kinesin-Related Proteins with a Mitochondrial Targeting Signal.";
 RL Plant Physiol. 127:724-726 (2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Hsuan V.W., Lee J.M.,
 RA Quach H.L., Tang C.C., Toriumi M., Wallender E.K., Wong C., Wu H.C.,
 RA Yu G., Yuan S., Chen H., Cheuk R., Jones T., Kim C.J., Nguyen M.,
 RA Palm C.J., Shinn P., Southwick A., Tripp M.G., Wu T.J., Davis R.W.,
 RA Ecker J.R., Theologis A.;
 RT "Arabidopsis Open Reading Frame (ORF) Clones.";
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB062739; BAB71852.1; -;
 DR EMBL; AY150516; AAN13032.1; -;
 DR GO; GO:0005871; C:kinesin complex; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0003774; F:motor activity; IEA.
 DR GO; GO:0007017; P:microtubule-based process; IEA.
 DR InterPro; IPR001752; kinesin_motor.
 DR Pfam; PF00225; kinesin_1.
 DR PRINTS; PR00380; KINESINHEAVY.
 DR SMART; SM00129; KISC; 1.
 DR SMART; SM00184; RING; 1.
 DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
 DR PROSITE; PS50067; KINESIN MOTOR_DOMAIN2; 1.
 DR PROSITE; PS50089; ZF_RING_2; 1.
 SQ SEQUENCE 1055 AA; 116463 MW; B69EC383FF14AB7B CRC64;

Query Match 39.3%; Score 686; DB 10; Length 1055;
 Best Local Similarity 46.2%; Pred. No. 2,6e-45;
 Matches 159; Conservative 57; Mismatches 114; Indels 14; Gaps 7;
 QY 1 AEGAVAVCVRPPLNSRESLGETAQVYKWTNNVIYQVSGKSNFDFVHGNETTGN 60
 Db 94 SERDSISVYRFRPLSDREYQGRQDEVAWYDGTLVREHYNPLTAVAFDVFQPGQATID 153
 QY 61 VYREIAAPIIDSALQGYNGTIFAYGOTASCKTYTMWSEHLGVIPRAIHDIQTKKFP 120
 Db 154 VYDVARPVVKAAGVNGTIFAYGVTSSGKTHTMGDQSPGIIPLAIKDVFSLIQDTP 213

QY 121 DREFLLRVSYMEIYNETITDILCGTQKMKPLIREDVNRNVYVADLTETEVYVSEMAKLM 180
 Db 214 GREFFLRVSYLRIYNEVINDLLDPTG--QNLRVRED--SQGTVEGKKEEVLSPGHLSF 270
 QY 181 IYKGSKRHYGTQKMNORSRSHITFRMILESGKEPNCESGVKVSHLNLDLAGSER 240
 Db 271 IAAGEHRHVGNFNLLSRSHITFTLMVESATGDEYD--GVTFSQLNLDLAGSE- 326
 QY 241 AAQTGAAGVRLKEGNCNINRSLFQVKKLSDGVGGFINTYRDSKLTILQNSLGGNPK 300
 Db 327 SSKTETGLRRKEGYSINKSLTLGTVICKLSEK-ATHIPYRDSKLTLLQSSLSGHGH 385
 QY 301 TRIICTITPV--SPDETITLQFASAK-----YMKNTPTVNEVS 338
 Db 386 VSLICTITPASSSSEETHNTLKFASRAKSIETIYASRNQIIDEKS 429

RESULT 13

QW5R5 PRELIMINARY; PRT; 699 AA.
 AC QW5R5;
 DT 01-OCT-2002 (TREMELrel. 22, Created)
 DT 01-OCT-2003 (TREMELrel. 25, Last sequence update)
 DE OSM-3 protein (corresponding sequence M02B7.3b).
 GN M02B7.3 OR OSM-3.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=Bristol N2;
 RX MEDLINE=99069613; PubMed=9851916;
 RA Waterston R.;
 RT "Genome sequence of the nematode C. elegans: a platform for
 RL investigating biology. The C. elegans Sequencing Consortium.";
 RN Science 282:2012-2018 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Nelson J., Wohldmann P.;
 RT "The sequence of C. elegans cosmid M02B7.";
 RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA STRAIN=Bristol N2;
 RA Waterston R.;
 RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U70853; AAM97997.1; -;
 DR Wormpep; M02B7.3b; CE31568.
 DR GO; GO:0005871; C:kinesin complex; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0003774; F:motor activity; IEA.
 DR GO; GO:0007017; P:microtubule-based process; IEA.
 DR InterPro; IPR001752; kinesin_motor.
 DR Pfam; PF00225; kinesin_1.
 DR PRINTS; PR00380; KINESINHEAVY.
 DR SMART; SM00129; KISC; 1.
 DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
 DR PROSITE; PS50067; KINESIN MOTOR_DOMAIN2; 1.
 SQ SEQUENCE 699 AA; 78779 MW; 8A774E3EF3A07813 CRC64;

Query Match 39.0%; Score 680; DB 5; Length 699;
 Best Local Similarity 46.4%; Pred. No. 4,2e-45;
 Matches 159; Conservative 53; Mismatches 109; Indels 22; Gaps 9;
 QY 5 AVAVCVRPPLNSRESLGETAQVYKWTNNVIYQV-----DG-SKSNFDFVHGNETT 58
 Db 4 SVRVAVRCRFPNQREKDLNTTLCV-GMTFN--VGQVNLNAPDGAAXDFTFDGAYFMDSTG 60
 QY 59 KNVYBIAPIIDSALQGYNGTIFAYGOTASCKTYTMWSEDL---HLGVIPRAIHDIQK 115

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 09:26:01 ; Search time 24.3176 Seconds
(without alignments)
4694.096 Million cell updates/sec

Title: US-10-045-631B-88_COPY_2_405

Perfect score: 2033

Sequence: 1 AEEGAVVCVRVPLNSREE.....LQKVQNEKIENLRMLVTSS 404

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04.*

- 1: Geneseqp1980s.*
- 2: Geneseqp1990s.*
- 3: Geneseqp2000s.*
- 4: Geneseqp2001s.*
- 5: Geneseqp2002s.*
- 6: Geneseqp2003as.*
- 7: Geneseqp2003bs.*
- 8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2053	100.0	2633	4	Abg06505 Novel hum
2	2053	100.0	2663	4	Aam39097 Human pol
3	1944.5	94.7	2688	4	Aam40883 Human pol
4	1468.5	71.5	2954	2	Aay01632 Amino aci
5	820	39.9	366	4	Abu53125 Intracell
6	781	38.0	348	4	Abu53208 Human cel
7	754	36.7	2013	4	Abb62322 Drosophil
8	704	34.3	677	4	Abb65183 Drosophil
9	694.5	33.8	473	5	Abg70992 Human tar
10	694.5	33.8	522	5	Abg70991 Human Hsk
11	694.5	33.8	1232	7	Add4938 Human lun
12	694.5	33.8	1232	7	Add18924 Human dis
13	694	33.8	1034	3	Aag31112 Arabidops
14	694	33.8	1069	3	Aag31111 Arabidops
15	694	33.8	1121	3	Aag31110 Arabidops
16	691.5	33.7	1029	7	Aae17786 Human kin
17	691.5	33.7	1029	7	Adc10190 Human NOV
18	684.5	33.3	1232	7	Add49337 Human kin
19	684.5	33.3	1232	7	Add49932 Human lun
20	683	33.3	1038	5	Aam48337 Human kin
21	682.5	33.2	1234	5	Abg70993 Human Hsk
22	678.5	33.0	1232	5	Abg70990 Human Hsk
23	662.5	32.3	1048	4	Abb59245 Human Hsk
24	662.5	32.3	1066	4	Aag67418 Drosophil
25	662.5	32.3	1066	6	Abg72693 Fruitfly

ALIGNMENTS

RESULT 1

ABG06505

ID ABG06505 standard; protein; 2633 AA.

XX AC ABG06505;

XX DT 13-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #6496.

XX KW Human; chromosome mapping; Gene mapping; Gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US008631.

XX PR 31-MAR-2000; 2000US-00540217.

XX PR 23-AUG-2000; 2000US-00649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Dmanac RT, Liu C, Tang YT;

XX XX WPI; 2001-639362/73.

XX DR N-PSDB; AAS70692.

XX XX New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity.

XX PS Claim 20; SEQ ID NO 36864; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying normal genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging

CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2633 AA;

Query Match 100.0%; Score 2053; DB 4; Length 2633;
Best Local Similarity 100.0%; Pred. No. 1.5e-171; Indels 0; Gaps 0;
Matches 404; Conservative 0; Mismatches 0;
QY 1 AEEGAVACVVRPLNSRESLSGETAQVYWKTDNNVIYQVDSKSFNDFRVEHNETTKN 60
DB 2 AEEGAVACVVRPLNSRESLSGETAQVYWKTDNNVIYQVDSKSFNDFRVEHNETTKN 61
QY 61 VYEIAPITDSAIQGVNGTIFAYGOTASGKTYTMGSEDLGVIPRAIHDFQIKKFP 120
DB 62 VYEIAPITDSAIQGVNGTIFAYGOTASGKTYTMGSEDLGVIPRAIHDFQIKKFP 121
QY 121 DREFLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNRYVADLTEEVVYVSEMAK 180
DB 122 DREFLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNRYVADLTEEVVYVSEMAK 181
QY 181 ITGKESRHYGETKMNQSRSHITFRMILESREKGEPSNCEGSKVSHNLVDLAGSER 240
DB 182 ITGKESRHYGETKMNQSRSHITFRMILESREKGEPSNCEGSKVSHNLVDLAGSER 241
QY 241 AAQTGAAGVRLKEGCNINRSLFILGVQVKKLSQGVGGFINYRDSKLTILQNSLGNPK 300
DB 242 AAQTGAAGVRLKEGCNINRSLFILGVQVKKLSQGVGGFINYRDSKLTILQNSLGNPK 301
QY 301 TRIICTITPVSFDETLTALQFASAKYMKNTPYVNEVSTDEALLKRYRKEIMDLKKOLEE 360
DB 302 TRIICTITPVSFDETLTALQFASAKYMKNTPYVNEVSTDEALLKRYRKEIMDLKKOLEE 361
QY 361 VSLKETRAQAMEKDQALQLEEKDLQKQVNEKIENLTMLVTSS 404
DB 362 VSLKETRAQAMEKDQALQLEEKDLQKQVNEKIENLTMLVTSS 405

RESULT 2
AAM39097
ID AAM39097 standard; protein; 2663 AA.
XX
AC AAM39097;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 2242.
XX
KW Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US034263.
XX
XX 23-DEC-1999; 99US-00471275.
PR

PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
PR 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
PR 29-NOV-2000; 2000US-00727344.
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
XX Zhou P, Goodrich R, Drmanac R;
XX
XX WPI; 2001-442253/47.
XX N-PSDB; AAI58253.
XX Novel nucleic acids and polypeptides, useful for treating disorders such
XX as central nervous system injuries.
XX
XX Example 4; SEQ ID NO 2242; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and the
XX encoded polypeptides (AAM38642-AAM42213) with nontropic,
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localised neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: Immune system suppression,
XX Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX assays for receptor activity, arthritis and inflammation, leukaemias and
XX C.N.S disorders. Note: The sequence data for this patent did not form
XX part of the printed specification
XX
SQ Sequence 2663 AA;

Query Match 100.0%; Score 2053; DB 4; Length 2663;
Best Local Similarity 100.0%; Pred. No. 1.5e-171;
Matches 404; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AEEGAVACVVRPLNSRESLSGETAQVYWKTDNNVIYQVDSKSFNDFRVEHNETTKN 60
DB 2 AEEGAVACVVRPLNSRESLSGETAQVYWKTDNNVIYQVDSKSFNDFRVEHNETTKN 61
QY 61 VYEIAPITDSAIQGVNGTIFAYGOTASGKTYTMGSEDLGVIPRAIHDFQIKKFP 120
DB 62 VYEIAPITDSAIQGVNGTIFAYGOTASGKTYTMGSEDLGVIPRAIHDFQIKKFP 121
QY 121 DREFLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNRYVADLTEEVVYVSEMAK 180
DB 122 DREFLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNRYVADLTEEVVYVSEMAK 181
QY 181 ITGKESRHYGETKMNQSRSHITFRMILESREKGEPSNCEGSKVSHNLVDLAGSER 240
DB 182 ITGKESRHYGETKMNQSRSHITFRMILESREKGEPSNCEGSKVSHNLVDLAGSER 241
QY 241 AAQTGAAGVRLKEGCNINRSLFILGVQVKKLSQGVGGFINYRDSKLTILQNSLGNPK 300
DB 242 AAQTGAAGVRLKEGCNINRSLFILGVQVKKLSQGVGGFINYRDSKLTILQNSLGNPK 301
QY 301 TRIICTITPVSFDETLTALQFASAKYMKNTPYVNEVSTDEALLKRYRKEIMDLKKOLEE 360
DB 302 TRIICTITPVSFDETLTALQFASAKYMKNTPYVNEVSTDEALLKRYRKEIMDLKKOLEE 361
QY 361 VSLKETRAQAMEKDQALQLEEKDLQKQVNEKIENLTMLVTSS 404
DB 362 VSLKETRAQAMEKDQALQLEEKDLQKQVNEKIENLTMLVTSS 405

RESULT 3
AAW40883
ID AAW40883 standard; protein; 2688 AA.
XX AC AAW40883;
XX DT 22-OCT-2001 (first entry)
XX DE Human polypeptide SEQ ID NO 5814.
XX KW Human; neutropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX KW peripheral nervous system; neuropathy; central nervous system; CNS;
XX KW Alzheimer's, Parkinson's disease; Huntington's disease; haemostatic;
XX KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX KW leukaemia.
XX OS Homo sapiens.
XX FN WO200153312-A1.
XX PD 26-JUL-2001.
XX PF 26-DEC-2000; 2000WO-US034263.
XX PR 23-DEC-1999; 99US-00471275.
XX PR 21-JAN-2000; 2000US-00488725.
XX PR 25-APR-2000; 2000US-0052317.
XX PR 20-JUN-2000; 2000US-00598042.
XX PR 19-JUL-2000; 2000US-00620312.
XX PR 03-AUG-2000; 2000US-00653450.
XX PR 14-SEP-2000; 2000US-00662191.
XX PR 19-OCT-2000; 2000US-00693036.
XX PR 29-NOV-2000; 2000US-00727344.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA,
XX PI Zhou P, Goodrich R, Drmanac RT;
XX PR WPI: 2001-442253/47.
XX PR N-PSDB; AAI60039.
XX PT Novel nucleic acids and polypeptides, useful for treating disorders such
XX PT as central nervous system injuries.
XX PS Example 2; SEQ ID NO 5814; 10078pp; English.
XX CC The invention relates to human nucleic acids (AAI57798-AAI61369) and the
XX CC encoded polypeptides (AAW38642-AAW42213) with neutropic,
XX CC immunosuppressant and cytostatic activity. The polynucleotides are useful
XX CC in gene therapy. A composition containing a polypeptide or polynucleotide
XX CC of the invention may be used to treat diseases of the peripheral nervous
XX CC system, such as peripheral nervous injuries, peripheral neuropathy and
XX CC localised neuropathies and central nervous system diseases, such as
XX CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX CC utilisation of the activities such as: Immune system suppression,
XX CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX CC assays for receptor activity, arthritis and inflammation, leukaemias and
XX CC C.N.S disorders. Note: The sequence data for this patent did not form
XX CC part of the printed specification
XX SQ Sequence 2688 AA;
XX
XX Query Match 94.7%; Score 1944.5; DB 4; Length 2688;
XX Best Local Similarity 96.1%; Pred. No. 6e-162;
XX Matches 391; Conservative 2; Mismatches 11; Indels 3; Gaps 3;
XX
XX 1 ABEQAVAVCVVRPLNSRRESLGETAQVYKTNHNNVIYQVDSKSFNDRVFGHNETTKN 60

Db 23 ABEQAVAVCVVRPLNSRRESLGETAQVYKTNHNNVIYQVDSKSFNDRVFGHNETTKN 82
QY 61 VYEETAAPIIDSAIQGYNGTIFAYGQTASGKTYTMMGSDHLGVIPRA-IHDFP-QKIKX 118
Db 83 VYEETAAPIIDSAIQGYNGTIFAYGQTASGKTYTMMGSDHLGVIPQGPHGHFSOKIXE 142
QY 119 -FPDRFLLRVSYMEIYNETITDLCGTQKMKPLIIRVDNRNRYVADLTVEEYVYSEMA 177
Db 143 VFDRFLLRVSYMEIYNETITDLCGTQKMKPLIIRVDNRNRYVADLTVEEYVYSEMA 202
QY 178 LKWITKGEKSRHYGETKMQRSRSHITPRMILESEKGEPCNCEGSKVSHNLVLDLAG 237
Db 203 LKWITKGEKSRHYGETKMQRSRSHITPRMILESEKGEPCNCEGSKVSHNLVLDLAG 262
QY 238 SRAAQTGAAGVRLKEGNCINRSLFILGOVKKLSGQGVGFNYRDSKLTILQNSLGG 297
Db 263 SRAAQTGAAGVRLKEGNCINRSLFILGOVKKLSGQGVGFNYRDSKLTILQNSLGG 322
QY 298 NPKTRILCIITPVSFDETLTALQFASTAKYMQNTPYVNEVSTDEALLKRYRKEIMDLKKQ 357
Db 323 NPKTRILCIITPVSFDETLTALQFASTAKYMQNTPYVNEVSTDEALLKRYRKEIMDLKKQ 382
QY 358 LEEVSLTETRAQAMEKDQLAQLLEEKDLLQKVQNEKIENLTMLVTSS 404
Db 383 LEEVSLTETRAQAMEKDQLAQLLEEKDLLQKVQNEKIENLTMLVTSS 429

RESULT 4
AAW01632
ID AAW01632 standard; protein; 2954 AA.
XX AC AAW01632;
XX DT 22-JUN-1999 (first entry)
XX DE Amino acid sequence of centromere-associated protein-E (CENP-E).
XX KW CENP-E; centromere-associated protein-E; ATPase activity;
XX KW plus end-directed microtubule motor activity; chromosome congression;
XX KW microtubule binding activity; chromosome movement; mitosis;
XX KW cell proliferation; tumor; metastasis; vascular malfunction;
XX KW inflammatory disease; immune disease; angiogenesis; hypertension;
XX KW restenosis; fungal infection; selective herbicide; fungicide;
XX KW insecticide; plant growth regulator; activator; cancer cell marker.
XX OS Xenopus sp.
XX PN WO9913061-A1.
XX PD 18-MAR-1999.
XX PF 10-SEP-1998; 98WO-US019231.
XX PR 11-SEP-1997; 97US-0058645P.
XX PA (REGC) UNIV CALIFORNIA.
XX PI Wood KW, Sakowicz R, Goldstein LSB, Cleveland DW;
XX PI WPI: 1999-229233/19.
XX PR N-PSDB; AAW26819.
XX PT Centromere-associated protein-E and related nucleic acid.
XX PS Claim 5; Page 66-67; 77pp; English.
XX CC The present sequence represents CENP-E (centromere-associated protein-E)
XX CC of Xenopus. The protein has at least one of plus end-directed microtubule
XX CC motor activity, ATPase (adenosine triphosphatase) activity and
XX CC microtubule binding activity. CENP-E is the motor that powers chromosome
XX CC movement toward microtubule plus ends and is essential for congression of
XX CC chromosomes during mitosis. Modulators of CENP-E can thus control cell

Nucleic acids having the sequences of clones isolated from libraries of different human tissues, useful in recombinant DNA methodologies.

Example III; Page 745; 1095pp; English.

This invention describes novel polynucleotides and polypeptides isolated from human cDNA libraries which can be used for gene therapy or in vaccines. The polynucleotides of the invention and antibodies encoded by them may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate polypeptide expression. The products of the CC invention may also be used to identify modulators of expression and CC activity and to down regulate expression and activity. The antibodies of CC the invention may also be used as diagnostic agents for detecting the CC presence of polypeptides in samples. This sequence represents a homologue CC of a polypeptide described in the disclosure of the invention XX
SQ Sequence 366 AA;

Query Match 39.9%; Score 820; DB 4; Length 366;
Best Local Similarity 50.5%; Pred. No. 9.8e-64;
Matches 187; Conservative 42; Mismatches 115; Indels 26; Gaps 7;

11 RVRPLNSREESGETAQVYV-----KTDNNVIYQVDSKSFNDRVEHGNETKNVYEE 64
1 RCRPLNREINDGSCVQWPPWTGYKTVHNGHSGSPHKSFTFDHVFVWNCQEDVYDT 60
65 IAAPITDSALQYNGTIFAYGQTASGKTYTMMG-----SEDLGVIPRAIHDFQIKKFP 120
61 VAHPYVDDCFHGYNCITIFAYGQTSGKTYTMMGPGGSEHPDMGLIPRCCHDIFDRDKFQ 120
121 DRG---FLLRVSWEIYNETITDLLC-GTQMKPLIREDVNRNRYVADLTEEVVYTSSEM 176
121 EKDHDFWHVKCSWEIYNEIYDLLCPNQHMKPLNIHEHPNMGPYVQGCTEFHVCSYED 180
177 ALKMITKGEKSRHYGETKMQRSRSHITIPRMILESEKGEPSNCEGSVKVSHLNLVDLA 236
181 ACHWIMQGNKNRVAATNMNDHSSRSHITIFTIHVEQRHK-----QCDEHVCHSKNVLVDLA 236
237 GSRAAQGTGAAGVRLKEGCNINRSFLTGQVVKLSGQV-----GGFINYRDSKLTTRI 290
237 GSERVNRGTAGGQRLKEGCNINRSFLTGQVVKLSGQV-----GGFINYRDSKLTWL 296
291 LQNSLGNPKTRIICITTPV--SFDETLTALQFASAKYMKNTPYNEVSTDEALLKRYR 348
297 LQDSLGNCKTCTIACIWPADNMYEETLTSLRYADRAKNKPKQINEDPCAMALWRYH 356
349 KEIMDLKKQL 358
357 EQIQDMKHQL 366

RESULT 6
ABU53208 standard; protein; 348 AA.

ABU53208 standard; protein; 348 AA.
XX AC ABU53208;
XX DT 14-APR-2003 (first entry)
XX DE Human cell cycle-associated DKFZp385b4 homologue #7.
XX KW Human; gene therapy; vaccine; disease treatment; detection.
XX OS Homo sapiens.

XX OS WO200112659-A2.
XX PN 22-FEB-2001.
XX PF 18-AUG-2000; 2000WO-IB001496.
XX PR 18-AUG-1999; 99US-0149499P.
XX PR 28-SEP-1999; 99US-0156503P.

proliferation. Agents that modulate CENP-E activity are lead therapeutic, CC bioagricultural and diagnostic agents, e.g. for treatment of unwanted CC cell proliferation (typical of many examples are tumors and metastases; CC vascular malfunction; inflammatory and immune diseases; angiogenesis; CC hypertension; restenosis; and fungal infections), also as plant- CC protection agents (selective herbicides, fungicides and insecticides) and CC plant growth regulators or activators for improving yields. CENP-E is CC also a diagnostic marker for dividing cells, including cancer cells XX
SQ Sequence 2954 AA;

Query Match 71.5%; Score 1468.5; DB 2; Length 2954;
Best Local Similarity 70.9%; Pred. No. 9e-120;
Matches 290; Conservative 52; Mismatches 60; Indels 7; Gaps 3;

1 AREGAVAVCVVRPLNSREESGETAQVYVKTNDNNVIYQVDSKSFNDRVEHGNETKN 60
2 SEGDAVKVCVRVPLIQREQ--GDQANLQWKAGNNTISQVDTGKSFNDRVFNHSESTSQ 59
61 VVEEIAAPITDSALQYNGTIFAYGQTASGKTYTMMGSEDLGVIPRAIHDFQIKKFP 120
60 IYQEIAPVIRLSALQYNGTIFAYGQTSSGKTYTMMGTPNSLGLIIPQAIQEVFKLIQELP 119
121 DREFLLRVSWIYNETITDLLCGTQMKPLIREDVNRNRYVADLTEEVVYTSSEMALKW 180
120 NREFLLRVSWIYNETVXDLLCDRRKKPLEIREDFNRNRYVADLTEELVMVPEHVIOW 179
181 ITKGEKSRHYGETKMQRSRSHITIPRMILESEKGEPS---NCEGSVKVSHLNLVDLAG 237
180 IKKGEKSRHYGETKMQRSRSHITIPRMILESEKGEPS---NCEGSVKVSHLNLVDLAG 239
238 SERAAQTGAAGVRLKEGCNINRSFLTGQVVKLSGQVGGFINYRDSKLTTRIQLNSLGG 297
240 SERASQTGAAGVRLKEGCNINRSFLTGQVVKLSGQVGGFINYRDSKLTTRIQLNSLGG 299
298 NPKTRIICITTPVSFDETLTALQFASAKYMKNTPYNEVSTDEALLKRYRKEIMDLKKQ 357
300 NAKTVIICITTPVSFDETLTALQFASAKYMKNTPYNEVSTDEALLKRYRKEIMDLKKQ 359
358 LE--EVSLETRAQAMKQDLAQLLEKDLQKQVNEKIENLTMVTS 404
360 LENLESSSETQAQAKEEHTQLLAETIKLHKEREDRIWHLTNIVVASS 408

RESULT 5
ABU53125 standard; protein; 366 AA.

ABU53125 standard; protein; 366 AA.
XX AC ABU53125;
XX DT 15-APR-2003 (first entry)
XX DE Intracellular trafficking-associated DKFZp385b3_26g22 homologue #2.
XX KW Human; gene therapy; vaccine; disease treatment; detection.

XX OS Homo sapiens.
XX PN WO200112659-A2.
XX PD 22-FEB-2001.
XX PF 18-AUG-2000; 2000WO-IB001496.

XX PR 18-AUG-1999; 99US-0149499P.
XX PR 28-SEP-1999; 99US-0156503P.
XX PA (GEHU-) GERMAN HUMAN GENOME PROJECT.
XX PI Wiemann S;
XX DR WPI; 2001-327840/34.
XX

XX PA (GEHU-) GERMAN HUMAN GENOME PROJECT.
 XX PI Wiemann S;
 XX XX WPI; 2001-327840/34.
 XX PT Nucleic acids having the sequences of clones isolated from libraries of
 XX PT different human tissues, useful in recombinant DNA methodologies.
 XX PS Example III; Page 829; 1095pp; English.
 XX CC This invention describes novel polynucleotides and polypeptides isolated
 CC from human cDNA libraries which can be used for gene therapy or in
 CC vaccines. The polynucleotides of the invention and antibodies encoded by
 CC them may be used in the prevention, diagnosis and treatment of diseases
 CC associated with inappropriate polypeptide expression. The products of the
 CC invention may also be used to identify modulators of expression and
 CC activity and to down regulate expression and activity. The antibodies of
 CC the invention may also be used as diagnostic agents for detecting the
 CC presence of polypeptides in samples. This sequence represents a homologue
 CC of a polypeptide described in the disclosure of the invention
 XX SQ Sequence 348 AA;
 Query Match 38.0%; Score 781; DB 4; Length 348;
 Best Local Similarity 51.1%; Pred. No. 2.6e-60;
 Matches 178; Conservative 38; Mismatches 106; Indels 26; Gaps 7;
 QY 11 RVRPLNSREESIGETAQVYVW-----KTDNNVIYQVDSKSFNDRVFGHNETTKNYYEE 64
 Db ||||| : : : : : ||| : : : : :
 QY 1 RCRPLNREINDGSCVQVQVPPWTGYKTVHNGHEGDSPHKSFTHDFVFWNCTQEDVYDT 60
 Db ||||| : : : : : ||| : : : : :
 QY 65 IAAPIIDSAIOGYNTIFAYGQTASGKTYTWMG-----SEDLHGVIPRAIHDFQKIKKFP 120
 Db ||||| : : : : : ||||| : : : : :
 QY 61 VAHPVDDCFHGYNTIFAYGQTSGKTYTWMPGGEHPDHWGIIPRCHDIFDRIDKFP 120
 Db ||||| : : : : : ||||| : : : : :
 QY 121 DRE--FLLRVSYMEIYNETITDILC-GTQMKPLIREDVNRNYYVADLTEEYVYVISEM 176
 Db : : : : : ||||| : : : : : ||||| : : : : :
 QY 121 EKDDHFWHVKSYMEIYNEETIYLLCPNPQHMPLNIHEHFMGYPYVQCTEFHVCVYED 180
 Db : : : : : ||||| : : : : : ||||| : : : : :
 QY 177 ALKWITGKSRHYGETKMNQSRSHITFRMILESREKGEPSNCEGSKVSHLNLVDLA 236
 Db ||||| : : : : : ||||| : : : : : ||||| : : : : :
 QY 181 ACHWIQGNKRNHVAATNMNDHRSRSHITFIHVEORHK-----QCDEHVCHSKNVLVDLA 236
 Db ||||| : : : : : ||||| : : : : : ||||| : : : : :
 QY 237 GSRAAQTAGVRLKEGCNINRSILFGLGVIKKLSGQV-----GSGFINYRDSKLTRI 290
 Db ||||| : : : : : ||||| : : : : : ||||| : : : : :
 QY 237 GSERVNTGAEGQRLKEGCNINQSLTLGNVINALADGQTKYMGHGHIPYRDSKLTWL 296
 Db ||||| : : : : : ||||| : : : : : ||||| : : : : :
 QY 291 LQNSLGNPKTRICTITPV--SPDETITLQFASTAKYMKNTPYVNE 336
 Db ||||| : : : : : ||||| : : : : : ||||| : : : : :
 QY 297 LQDSLGNCKTMIACIWPADWNYEETLSTLRVADRAKNKPKQINE 344
 Db ||||| : : : : : ||||| : : : : : ||||| : : : : :
 RESULT 7
 ABB62322
 ID ABB62322 standard; protein; 2013 AA.
 XX AC ABB62322;
 XX AC ABB62322;
 XX DT 26-MAR-2002 (first entry)
 XX DE Drosophila melanogaster polypeptide SEQ ID NO 13758.
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 OS Drosophila melanogaster.
 XX FN W0200171042-A2.
 XX PD 27-SEP-2001.
 XX DE Drosophila melanogaster polypeptide SEQ ID NO 22341.

PF 23-MAR-2001; 2001WO-US009231.
 XX 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX (PEKE) PE CORP NY.
 XX Venter JC, Adams M, Li PWD, Myers EW;
 XX WPI; 2001-656860/75.
 DR N-PSDB; ABL06425.
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.
 XX Disclosure; SEQ ID NO 13758; 21pp + Sequence Listing; English.
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX SQ Sequence 2013 AA;
 Query Match 36.7%; Score 754; DB 4; Length 2013;
 Best Local Similarity 45.4%; Pred. No. 8.4e-57;
 Matches 181; Conservative 62; Mismatches 118; Indels 38; Gaps 10;
 QY 5 AVAVCVVRPLNSREESIGETAQVYVYVKT-DNNVIYQVDSKSFNDRVFGHNETTKNYY 62
 Db : : : : : ||||| : : : : : ||||| : : : : :
 QY 8 SIQVICVRPCEPGLTSL-----WQKERRSIHLADSHAEPYVDFYVDEGASNOEVF 60
 Db : : : : : ||||| : : : : : ||||| : : : : :
 QY 63 BEIAAPIIDSAIOGYNTIFAYGQTASGKTYTWMGSDHLGVIIPRAIHDFQKIKKFPDR 122
 Db : : : : : ||||| : : : : : ||||| : : : : :
 QY 61 DRMAKHIVHACMQGFGNGTIFAYGQTSSGKTYTWMGDEQNPGVWVLAKEIFQQISSETER 120
 Db : : : : : ||||| : : : : : ||||| : : : : :
 QY 123 EFLRVSYMEIYNETITDILCGTQMKPLIREDVNRNYYVADLTEEYVYVISEM-ALKWI 181
 Db ||||| : : : : : ||||| : : : : : ||||| : : : : :
 QY 121 DFLRVGYIIEYNEKIYDILL--NKNQDLKTHESGNGIVNVN--CECIIITSEVDLRL 176
 Db ||||| : : : : : ||||| : : : : : ||||| : : : : :
 QY 182 TKGKSRHYGETKMNQSRSHITFRMILESREKGEPSNCEGSKVSHLNLVDLAGSERA 241
 Db ||||| : : : : : ||||| : : : : : ||||| : : : : :
 QY 177 CLGNKERTVGETNWNERSRSHALFKIIEESR-KSDHSD-DDAVIQSVLNLVDLAGSERA 234
 Db ||||| : : : : : ||||| : : : : : ||||| : : : : :
 QY 242 AQTGAAGVRLKEGCNINRSILFGLGVIKKLSGQVGFNVRDSKLTTRILQNSLGNPKPT 301
 Db ||||| : : : : : ||||| : : : : : ||||| : : : : :
 QY 235 DQTGARGARLKEGHINKSLFLSNVKSISENADNRFNYRDSKLTTRILQASLGNAPT 294
 Db ||||| : : : : : ||||| : : : : : ||||| : : : : :
 QY 302 RIICITTPSFDITLQFASTAKYMKNTPYVNEVSTDEALKRYRKEIMDLKQLEEV 361
 Db ||||| : : : : : ||||| : : : : : ||||| : : : : :
 QY 295 SIICITKPSIMEESQSTLSFATRAKKIRIKPQVNMVSDATMMKRLEREIKVLKDLAE- 353
 QY 362 SLETRAQAMEKQDLQALLEKDLQKQVNEKIENTLRL 400
 Db ||||| : : : : : ||||| : : : : : ||||| : : : : :
 QY 354 -----BER-----KNENQOKVEHLERQI 371
 Db ||||| : : : : : ||||| : : : : : ||||| : : : : :
 RESULT 8
 ABB65183
 ID ABB65183 standard; protein; 677 AA.
 XX AC ABB65183;
 XX AC ABB65183;
 XX DT 26-MAR-2002 (first entry)
 XX DE Drosophila melanogaster polypeptide SEQ ID NO 22341.

XX Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL09286.

XX New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions.

XX Disclosure; SEQ ID NO 22341; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA sequences (ABU1840-ABU16175) and the encoded proteins (ABU57737-ABU72072). The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 677 AA;

Query Match 34.3%; Score 704; DB 4; Length 677;

Best Local Similarity 43.0%; Pred. No. 4.4e-53;

Matches 176; Conservative 67; Mismatches 138; Indels 28; Gaps 11;

QY 6 VAVCVVRPLNSREESLG-ETAQVYWKTDNNVIYQV-----DGSKFNDVFGHNETT 58

DB 21 VRVVVTRPMDKNLSAGLSAISVDKINRAITVMKPNATANEPPKTYFDNFDGGSNQ 80

QY 59 KNYVEETAAPIIDSAIOGYNCTIFAYGOTASGKTYTMMGSDH---LGVIPRAIHDPQK 115

DB 81 MDLYVDTARPIVDKVLGNGTILAYGOTGKTYTMSGNDSPQTKGIIPNAFAHIFGH 140

QY 116 IKKPPD-REFLLRVSMEIYNETITDLCGTQKMKPLIREDVNRNVVADLTREVVYTS 174

DB 141 IAKAKENQKFLVRVSMEIYNEEVRDLL-GKDVGKSLVKERPDIGVFKDLGSLGVYHNA 199

QY 175 EMALKWITKGEKSRHYGTOMQRSRSHHTIFRMILESEKPEPNCESGSKVKSHLNLDV 234

DB 200 DDLENIMRLGNKRAVAGATKKNQESRSHAFISITVERSELGEGD--VOHVRMGKLQILD 257

QY 235 LAGSRAAQGTAGVRLKGCNINPSFLTGQVKKLSDGVGVGFINVRDSKLTIRLQNS 294

DB 258 LAGSRQKTQASGQRKKAATKINLSLVGNVIALVDGK-STHIFRNSKLTIRLLQDS 316

QY 295 LGGNKPKTRICITTPV--SFDETLTALQFASAKYMKNTPPYNEVSTDEALLKRYRKEIM 352

DB 317 LGGNSKTYWCATISPADSNYMETISTLRVYSAKNIQNRMHNEEPKD-ALLRHFQEETA 375

QY 353 DLKKOLEE-VSTETRAQAMEK-----DQLAQLE---EKDLQKQVONE 391

DB 376 RLKQLEEGDSEEPSPSEEBEDTADDEAPLEIELESSTIOAVEKK 424

RESULT 9

ABG70992

ID ABG70992 standard; protein; 473 AA.

XX AC ABG70992;

XX 10-DEC-2002 (first entry)

XX Human target protein.

XX Microtubule stimulated ATPase; cellular proliferation; cancer; AIDS; hyperplasia; restenosis; cardiac hypertrophy; immune disorder; HIV; inflammation; autoimmune disease; arthritis; graft rejection; psoriasis; inflammatory bowel disease; proliferation; medical procedure; surgery; human immunodeficiency virus; acquired immunodeficiency syndrome; angioplasty; human; HsKif; kinesin family.

XX Homo sapiens.

XX US6440694-B1.

XX 27-AUG-2002.

XX 12-JUN-2000; 2000US-00592054.

XX 12-JUN-2000; 2000US-00592054.

XX (CYTO-) CYTOKINETICS INC.

XX Beraud C, Finer JT, Sakowicz R, Wood KW;

XX WPI; 2002-711529/77.

XX N-PSDB; ABS55162.

XX Screening for modulators of target protein having microtubule stimulated ATPase activity e.g. kinesin family of protein, useful for treating cancer, psoriasis, arthritis, human immunodeficiency virus (HIV) infection.

XX Claim 3; Fig 6; 34pp; English.

XX The present invention relates to a new method of screening modulators of target protein with microtubule stimulated ATPase activity. The method involves contacting the target protein with an agent at 1st and 2nd concentrations and determining the level of activity (e.g. binding or ATPase activity) of target protein, where a difference between levels of activity of target protein contacted with 1st and 2nd concentrations of an agent indicates that an agent modulates activity of target protein. The invention can be used for screening for modulators of target protein having microtubule stimulated ATPase activity. The compounds identified by method of the invention are useful for treating cellular proliferation including cancer, hyperplasias, restenosis, cardiac hypertrophy, immune disorders and inflammation. The compounds identified by the method are also useful for treating autoimmune disease, arthritis, graft rejection, inflammatory bowel disease, proliferation induced by medical procedures, e.g. surgery, angioplasty etc. The compounds are also useful for treating psoriasis. The compounds are useful for inhibiting human immunodeficiency virus (HIV) and thus treating acquired immunodeficiency syndrome (AIDS). The present amino acid sequence represents the human target protein of the invention

XX Sequence 473 AA;

Query Match 33.8%; Score 694.5; DB 5; Length 473;

Best Local Similarity 39.6%; Pred. No. 1.8e-52;

Matches 168; Conservative 66; Mismatches 145; Indels 45; Gaps 8;

QY 6 VAVCVVRPLNSREESLG-ETAQVYWKTDNNVIYQVDSKSFNDFRVEHNETTKNVYEE 64

DB 8 VRVALRCRLVPKEISEGCMCLSFVPGEPQVVVGTD--KSFTYDFVDFPSTQEVEFNT 65

QY 65 IAAPIIDSAIOGYNCTIFAYGOTASGKTYTMMG-----SEDLGVIPRAIHDPQKIK 117

Db 66 AVAPLIKGVFGYNATLAYQGTSGKTSYMGAYTAEQNEPTGVIPRVQLLFKEID 125
 Qy 118 KFPDREFLLRVSYMEIYNETITDILCGTQKMKPLIREDVNRNVVADLTREVVYTSEMA 177
 Db 126 KKSDFEFLKVSYLEIYNEEILDLCPREKAQINIREDPKEGKIVGLTEKTVLVALDT 185
 Qy 178 LKWITKGEKSRHYGETKMNORSRSHITFRMILIESREKGEPSNCEGSKVSHLNLVDLAG 237
 Db 186 VSCLEQGNNSRTVASTAMNSQSSRSHAFITISLEQRKSD----KNSFRSKLHLVDLAG 241
 Qy 238 SERAAQTGAAGVRLKEGCNINRSLFILGQVVKLSGQGVGFNYRDSKLTIRLQNSLGG 297
 Db 242 SERQKTKAEGDRLEKGININRGLCLGNVISALGDDKGGFVPRYRDSKLTIRLQNSLGG 301
 Qy 298 NPKTRIICTITPV--SFDETLTALQFSTAKYMKNTVPVNEVSTDEALLKRYKEIMDLK 355
 Db 302 NSHTLMACVSPADSNLEETILTRYADRARKINKKPIVN-----IDPQTAEHLNHLK 353
 Qy 356 KQLEEVSL-----ETRAQAMEKQDLAQLEEKDLQKQVNEKIENL 396
 Db 354 QVQVQLQVLLQLQAHGGTLPGSITVEPSENLSLMEKNQ--SLVEENEKLSRGLSEAAQT 411
 Qy 397 TRML 400
 Db 412 AQML 415

RESULT 10
 ABG70991
 ID ABG70991 standard; protein; 522 AA.
 AC
 XX
 XX
 DT 10-DEC-2002 (first entry)
 DE
 DE
 XX
 KW Microtubule stimulated ATPase; cellular proliferation; cancer; AIDS;
 KW hyperplasia; restenosis; cardiac hypertrophy; immune disorder; HIV;
 KW inflammation; autoimmune disease; arthritis; graft rejection; psoriasis;
 KW inflammatory bowel disease; proliferation; medical procedure; surgery;
 KW human immunodeficiency virus; acquired immunodeficiency syndrome;
 KW angioplasty; human; Hskif4; kinesin family.
 XX
 OS Homo sapiens.
 FH
 FT Key Location/Qualifiers
 FT Region 1..22
 FT /note= "N-terminal T7 epitope"
 FT Region 496..516
 FT /note= "C-terminal myc epitope"
 FT Region 517..522
 FT /note= "6-histidine residues at C-terminus"
 FT
 XX US6440684-B1.
 XX
 XX
 XX 27-AUG-2002.
 XX
 XX 12-JUN-2000; 2000US-00592054.
 XX
 XX 12-JUN-2000; 2000US-00592054.
 XX
 XX (CYTO-) CYTOKINETICS INC.
 XX
 XX Beraud C, Finer JT, Sakowicz R, Wood KW;
 XX
 XX WPI; 2002-711529/77.
 XX N-PSDB; ABS55161.
 XX
 XX Screening for modulators of target protein having microtubule stimulated
 XX ATPase activity e.g. kinesin family of protein, useful for treating
 XX cancer, psoriasis, arthritis, human immunodeficiency virus (HIV)

PT infection.
 XX Claim 2; Fig 4; 34pp; English.
 PS
 XX The present invention relates to a new method of screening modulators of
 CC target protein with microtubule stimulated ATPase activity. The method
 CC involves contacting the target protein with an agent at 1st and 2nd
 CC concentrations and determining the level of activity (e.g. binding or
 CC ATPase activity) of target protein, where a difference between levels of
 CC activity of target protein contacted with 1st and 2nd concentrations of
 CC an agent indicates that an agent modulates activity of target protein.
 CC The invention can be used for screening for modulators of target protein
 CC having microtubule stimulated ATPase activity. The compounds identified
 CC by method of the invention are useful for treating cellular proliferation
 CC including cancer, hyperplasias, restenosis, cardiac hypertrophy, immune
 CC disorders and inflammation. The compounds identified by the method are
 CC also useful for treating autoimmune disease, arthritis, graft rejection,
 CC inflammatory bowel disease, proliferation induced by medical procedures,
 CC e.g. surgery, angioplasty etc. The compounds are also useful for treating
 CC psoriasis. The compounds are useful for inhibiting human immunodeficiency
 CC virus (HIV) and thus treating acquired immunodeficiency syndrome (AIDS).
 CC The present amino acid sequence represents the human Hskif4 (kinesin
 CC family) construct protein of the invention
 XX
 SQ Sequence 522 AA;

Query Match 33.8%; Score 694.5; DB 5; Length 522;
 Best Local Similarity 39.6%; Pred. No. 2.1e-52;
 Matches 168; Conservative 66; Mismatches 145; Indels 45; Gaps 8;
 Qy 6 VAVCVRVPLNSREESLG-ETAQVYMTDNVYQVDSKSFNDFRVFHGNETTKNVVEE 64
 Db 30 VRVALRCPLVPKPEISEGQCMCLSFVPEGEQVVGTD--KSFYTFVDFPSTGEVEVNT 87
 Qy 65 IAAPIDSAIQYNGTTFPAYQGTASGKTYTMG-----SEHLGVIPRAIHDFQIK 117
 Db 88 AVAPLIKGVFGYNATLAYQGTSGKTSYMGAYTAEQNEPTGVIPRVQLLFKEID 147
 Qy 118 KFPDREFLLRVSYMEIYNETITDILCGTQKMKPLIREDVNRNVVADLTREVVYTSEMA 177
 Db 148 KKSDFEFLKVSYLEIYNEEILDLCPREKAQINIREDPKEGKIVGLTEKTVLVALDT 207
 Qy 178 LKWITKGEKSRHYGETKMNORSRSHITFRMILIESREKGEPSNCEGSKVSHLNLVDLAG 237
 Db 208 VSCLEQGNNSRTVASTAMNSQSSRSHAFITISLEQRKSD----KNSFRSKLHLVDLAG 263
 Qy 238 SERAAQTGAAGVRLKEGCNINRSLFILGQVVKLSGQGVGFNYRDSKLTIRLQNSLGG 297
 Db 264 SERQKTKAEGDRLEKGININRGLCLGNVISALGDDKGGFVPRYRDSKLTIRLQNSLGG 323
 Qy 298 NPKTRIICTITPV--SFDETLTALQFSTAKYMKNTVPVNEVSTDEALLKRYKEIMDLK 355
 Db 324 NSHTLMACVSPADSNLEETILTRYADRARKINKKPIVN-----IDPQTAEHLNHLK 375
 Qy 356 KQLEEVSL-----ETRAQAMEKQDLAQLEEKDLQKQVNEKIENL 396
 Db 376 QVQVQLQVLLQLQAHGGTLPGSITVEPSENLSLMEKNQ--SLVEENEKLSRGLSEAAQT 433
 Qy 397 TRML 400
 Db 434 AQML 437

RESULT 11
 ADD49938
 ID ADD49938 standard; protein; 1232 AA.
 XX
 AC ADD49938;
 XX
 DT 15-JAN-2004 (first entry)
 XX
 DE Human lung specific tumour antigen L 1447p.
 XX

KW Human; lung cancer antigen; cytostatic; lung cancer; gene therapy;
 KM vaccine; T-cell; tumour.
 XX Homo sapiens.
 OS US2003194764-A1.
 PN 16-OCT-2003.
 PD 04-APR-2002; 2002US-00116712.
 PF 05-APR-2001; 2001US-0282289P.
 PR 05-OCT-2001; 2001US-0327511P.
 XX (CORI-) CORIXA CORP.
 PA Bangur CS, Switzer A;
 PI WPI; 2003-844452/78.
 DR N-PSDB; ADD49936.
 DR New isolated polypeptides and polynucleotides useful for diagnosing,
 PT preventing and treating cancer, particularly lung cancer.
 PS Claim 1; SEQ ID NO 670; 250pp; English.
 XX The invention relates to an isolated polynucleotide (a) comprising any of
 CC the 666 fully defined nucleotide sequences appearing as ADD49269 -
 CC ADD49936, ADD49936 and ADD49938, complements of (a); sequences of at
 CC least 20 contiguous residues of (a); sequences that hybridise to (a)
 CC under highly stringent conditions; sequences having at least 75 or 90%
 CC identity to (a); or degenerate variants of (a). Also included are an
 CC isolated polypeptide (b) (comprising: sequences encoded by the new
 CC polynucleotide; any of the 4 amino acid sequences fully defined in the
 CC polynucleotide; or sequences having at least 70 or 90% identity to the
 CC sequence in (a) or (b)), an expression vector comprising the above,
 CC a polynucleotide operably linked to an expression control sequence, a host
 CC cell transformed or transfected with the above expression vector, an
 CC isolated antibody, or its antigen-binding fragment, that specifically
 CC binds to the above polypeptide, an oligonucleotide that hybridises to
 CC above-mentioned nucleotide sequences under highly stringent conditions,
 CC a fusion protein comprising at least one polypeptide cited above, detecting
 CC the presence of a cancer in a patient (comprising: obtaining a biological
 CC sample from the patient; contacting the biological sample with a binding
 CC agent that binds to the polypeptide, or with the oligonucleotide cited
 CC above; detecting in the sample an amount of the polypeptide that binds to
 CC the binding agent, or an amount of a polynucleotide that hybridises to
 CC the oligonucleotide; and comparing the amount of polypeptide, or
 CC polynucleotide that hybridises to the oligonucleotide, to a predetermined
 CC cut-off value and then determining the presence of a cancer in the
 CC patient), a method for stimulating and/or expanding T-cells specific for
 CC a tumour protein (comprising contacting T-cells with the above
 CC polypeptide, polynucleotide or antigen-presenting cells that express the
 CC polynucleotide, under conditions and for a time sufficient to permit the
 CC stimulation and/or expansion of T-cells), an isolated T-cell population
 CC comprising T-cells prepared by the method, a composition comprising a
 CC first component selected from physiological carriers and
 CC immunostimulants, and a second component selected from the above
 CC polypeptide, polynucleotide, antibody, fusion protein, T-cell population
 CC and antigen-presenting cells that express the above polypeptide,
 CC stimulating an immune response in a patient (comprising administering to
 CC the patient the above composition) treating lung cancer in a patient
 CC (comprising administering to the patient the above composition and a
 CC diagnostic kit (comprising: at least one oligonucleotide cited above; or
 CC at least one antibody cited above and a detection reagent, where the
 CC detection reagent comprises a reporter group). The composition and
 CC methods are useful in diagnosing, preventing and treating cancer,
 CC particularly lung cancer. The present sequence is a lung cancer-
 CC associated antigen of the invention.

SQ Sequence 1232 AA;

Query Match 33.8%; Score 694.5; DB 7; Length 1232;

Best Local Similarity 39.6%; Pred. No. 7.5e-52; Matches 168; Conservative 66; Mismatches 145; Indels 45; Gaps 8;	
QY	6 VAVCVVRPLNSREESLG-ETAQVYWKTDNNVYVQDGSKSNFDRVPHGNETTTNNVVE 64
DB	10 VRVALRCRPLVPKEISEGQMCLSFVGPBPQVVGTD--KSFTYDFVDPSTEQBEVENT 67
QY	65 IAAPIDSIAOIGYNGTIFAYGQTASGKYTYMG-----SEDLGVIPRAHDFQKIK 117
DB	68 AVAPLIKGVFGYNATVLAAYGQTSGKTYSMGGAYTAQENEPYGVIPRVQLLFKEID 127
QY	118 KPDEFEFLRYSYMEIYNETIDLLCGTKMKPLIIREDVNRNVVADLTESVWVTSEMA 177
DB	128 KKSDEFTLKVSLEYINEEILLDCPSREKQAINREDPKGKIKVGLTEKTVLVALDT 187
QY	178 LKWITKGEKSRHYGETKMNQSRSHITFRMILESREKGEPCNCGSVKSHNLVDLAG 237
DB	198 VSCLEGGNNRTVASTAMNSQSSRSHAITFTISLEQRKKS-----KNSSFRSKHLVDLAG 243
QY	238 SERAAQTGAAGVRLKEGNCNINRSILFQVINKLSDGVGGFNYRDSKLTILQNSLGG 297
DB	244 SERQKTKAEGDRLXEGININRGLLCLGNVISALGDDKGGFVYRDSKLTILQNSLGG 303
QY	298 NPKTRILICTITPV--SFDETLTALOFASTAKYMKNTPYVNEVSTDEALKRYRKEIMDLK 355
DB	304 NHTLMIACVSPADSNLEETNLRYADPARKIKKPIV-----IDPQTAEHLHLK 355
QY	356 KQLEEVSL-----ETRAQAMKQQAQLLEKDLQKQVONEKIENL 396
DB	356 QOVQQLQVLLQAHGQTLPGSITVPESENQLSMEKNQ--SLVEENEKLSRGLSEAAQGT 413
QY	397 TRML 400
DB	414 AQML 417
RESULT 12	
ID	ADD18924
ID	ADD18924 standard; protein; 1232 AA.
XX	ADD18924;
XX	15-JAN-2004 (first entry)
DE	Human disease related protein SeqID413.
XX	human; disease state; cytostatic; antiinflammatory; ophthalmological; antiarteriosclerotic; vulnery; gene therapy; hypoxia-regulated condition; tumorigenesis; angiogenesis; apoptosis; inflammation; erythropoiesis; glycolysis; gluconeogenesis; glucose transport; catecholamine synthesis; iron transport; nitric oxide synthesis; cancer; ischaemic condition; reperfusion injury; retinopathy; neonatal stress; pre-eclampsia; atherosclerosis; inflammatory condition; wound healing.
OS	Homo sapiens.
XX	WC2003018621-A2.
XX	06-MAR-2003.
XX	23-AUG-2002; 2002WO-GB003892.
XX	23-AUG-2001; 2001GB-00020558.
XX	05-OCT-2001; 2001GB-00024037.
XX	(OXFO-) OXFORD BIOMEDICA UK LTD.
XX	Kingman SM, White J, Ward NR, Harris RA, Naylor S, Mundy CR;
XX	WPI; 2003-290046/28.
XX	N-PSDB; ADD18925.

PT New substantially purified polypeptide, useful for diagnosing or treating
 PT a hypoxia-regulated condition, such as cancer, ischemia, reperfusion
 PT injury, retinopathy, pre-eclampsia, atherosclerosis, inflammation, or
 PT wound healing.

XX Claim 25; SEQ ID NO 413; 424pp; English.

XX PS
 XX PS
 CC This invention relates to novel human genes and gene product which are
 CC implicated in certain disease states. Compounds which modulate the
 CC proteins of the invention may have cytostatic, antiinflammatory,
 CC ophthalmological, antiarteriosclerotic or vulnerary activities. The
 CC sequences of the invention may be useful for gene therapy. The invention
 CC may be useful for diagnosing or treating a hypoxia-regulated condition,
 CC such as tumorigenesis, angiogenesis, apoptosis, inflammation,
 CC erythropoiesis, or the biological response to hypoxia conditions
 CC including processes such as glycolysis, gluconeogenesis, glucose
 CC transportation, catecholamine synthesis, iron transport or nitric oxide
 CC synthesis. The disease includes cancer, ischaemic conditions, reperfusion
 CC injury, retinopathy, neonatal stress, pre-eclampsia, atherosclerosis,
 CC inflammatory conditions or wound healing. The present sequence is that of
 CC a disease related protein of the invention.

XX Sequence 1232 AA;

Query Match 33.8%; Score 694.5; DB 7; Length 1232;

Best Local Similarity 39.6%; Pred. No. 7.5e-52;

Matches 168; Conservative 66; Mismatches 145; Indels 45; Gaps 8;

QY 6 VAVCVVRPLNSRBSLG-ETAQVWTKTDNNVIYQVDSKSNFDPVPHGNETTKNVEE 64
 Db 10 VRVALRCEPLVPKSEIGCQCLSFVPGEPQVVGTGDKSFTYDFVDPSTEQEEVENT 67
 QY 65 IAAPIIDSAIQYNGTIPAYQGTASGKTYTWG-----SEDLGVIPRAIHDFOKIK 117
 Db 68 AVAPLIXGVFKGNATVLAYGTGSGKTYSMGAYTAQENEPVGVIPVLIQLFREID 127
 QY 118 KFPDPRELLRYSMEIYNETTDLICGTQRMKPLIREDVNRNVVADLVEEVVYTSMA 177
 Db 128 KKSDFEFLKVSYLEIYNEEILDLCPSSREKAQINREDPKGKIVGLTEKTVLVALDT 187
 QY 178 LKWTGKESRHYGETKMNQSSSHTIFRMILSREKGEPSNCEGSKVSHMLVLDLAG 237
 Db 188 VSCLEQGNSTVASTANSSSHAFITLSLQKSD-----KNSFSRSLHLVLDLAG 243
 QY 238 SERAAQTGAAGVRKEGCNINRSFLIGQVTKLSGQGVGFVNRVSKLTRILQNSLGG 297
 Db 244 SERQKTKAEGDRLEKGINRGLGNGVLSALGDDKGGFVYRVSRLTRILQDSLGG 303
 QY 298 NPKTRIICITPV--SFDEITLQFASAKYMTQVYNEVSTDEALLKRYRKEIMDLK 355
 Db 304 NSHTLMACVSPADSNLEETLNTLYADRAKIKNKEIVN-----IDPQTAELNHLK 355
 QY 356 KOLEEVSLSL-----ETRAQAMEKDLQALKEEKDLQKQVNEKLENL 396
 Db 356 QVQVQLQVLLQAHGGTLPGSITVPSSENLSQSMKQ--SLVEENKSLRGLSEAAQGT 413
 QY 397 TRML 400
 Db 414 AQML 417

RESULT 13

AAG31112

ID AAG31112 standard; protein; 1034 AA.

XX AC

XX AAG31112;

XX DT

DT 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 37309.

DE DE

XX Protein identification; signal transduction pathway; metabolic pathway;

XX KW

XX hybridisation assay; genetic mapping; gene expression control; promoter;

KW

KW termination sequence.
 XX Arabidopsis thaliana.
 OS Arabidopsis thaliana.
 XX EP1033405-A2.
 XX XX
 PD 06-SEP-2000.
 XX XX
 PF 25-FEB-2000; 2000EP-00301439.
 XX 25-FEB-1999; 99US-0121825P.
 PR 05-MAR-1999; 99US-0123180P.
 PR 09-MAR-1999; 99US-0123548P.
 PR 23-MAR-1999; 99US-0125788P.
 PR 25-MAR-1999; 99US-0126264P.
 PR 29-MAR-1999; 99US-0126785P.
 PR 01-APR-1999; 99US-0127462P.
 PR 06-APR-1999; 99US-0128234P.
 PR 08-APR-1999; 99US-0128714P.
 PR 16-APR-1999; 99US-0129845P.
 PR 19-APR-1999; 99US-0130077P.
 PR 21-APR-1999; 99US-0130449P.
 PR 23-APR-1999; 99US-0130510P.
 PR 23-APR-1999; 99US-0130891P.
 PR 28-APR-1999; 99US-0131449P.
 PR 30-APR-1999; 99US-0132048P.
 PR 30-APR-1999; 99US-0132407P.
 PR 04-MAY-1999; 99US-0132484P.
 PR 05-MAY-1999; 99US-0132485P.
 PR 06-MAY-1999; 99US-0132486P.
 PR 06-MAY-1999; 99US-0132487P.
 PR 07-MAY-1999; 99US-0132863P.
 PR 11-MAY-1999; 99US-0134256P.
 PR 14-MAY-1999; 99US-0134218P.
 PR 14-MAY-1999; 99US-0134219P.
 PR 14-MAY-1999; 99US-0134221P.
 PR 14-MAY-1999; 99US-0134370P.
 PR 18-MAY-1999; 99US-0134768P.
 PR 19-MAY-1999; 99US-0134941P.
 PR 20-MAY-1999; 99US-0135124P.
 PR 21-MAY-1999; 99US-0135353P.
 PR 24-MAY-1999; 99US-0135629P.
 PR 25-MAY-1999; 99US-0136021P.
 PR 27-MAY-1999; 99US-0136392P.
 PR 28-MAY-1999; 99US-0136782P.
 PR 01-JUN-1999; 99US-0137222P.
 PR 03-JUN-1999; 99US-0137528P.
 PR 04-JUN-1999; 99US-0137502P.
 PR 07-JUN-1999; 99US-0137724P.
 PR 08-JUN-1999; 99US-0138094P.
 PR 10-JUN-1999; 99US-0138540P.
 PR 10-JUN-1999; 99US-0138847P.
 PR 14-JUN-1999; 99US-0139119P.
 PR 16-JUN-1999; 99US-0139452P.
 PR 16-JUN-1999; 99US-0139453P.
 PR 17-JUN-1999; 99US-0139492P.
 PR 18-JUN-1999; 99US-0139454P.
 PR 18-JUN-1999; 99US-0139455P.
 PR 18-JUN-1999; 99US-0139456P.
 PR 18-JUN-1999; 99US-0139457P.
 PR 18-JUN-1999; 99US-0139458P.
 PR 18-JUN-1999; 99US-0139459P.
 PR 18-JUN-1999; 99US-0139460P.
 PR 18-JUN-1999; 99US-0139461P.
 PR 18-JUN-1999; 99US-0139462P.
 PR 18-JUN-1999; 99US-0139463P.
 PR 18-JUN-1999; 99US-0139750P.
 PR 18-JUN-1999; 99US-0139763P.
 PR 21-JUN-1999; 99US-0139817P.
 PR 22-JUN-1999; 99US-0139899P.
 PR 23-JUN-1999; 99US-0140353P.
 PR 23-JUN-1999; 99US-0140354P.
 PR 24-JUN-1999; 99US-0140695P.

PR	28-JUN-1999;	99US-0140823P.	PR	10-SEP-1999;	99US-0153070P.	Query Match	33.8%;	Score 694;	DB 3;	Length 1034;
PR	29-JUN-1999;	99US-0140991P.	PR	13-SEP-1999;	99US-0153758P.	Best Local Similarity	36.2%;	Pred. No. 6.4e-52;		
PR	30-JUN-1999;	99US-0141287P.	PR	15-SEP-1999;	99US-0154018P.	Matches 178;	Conservative	87;	Mismatches 131;	Indels 96;
PR	01-JUL-1999;	99US-0141842P.	PR	16-SEP-1999;	99US-0154039P.					Gaps 14;
PR	01-JUL-1999;	99US-0142154P.	PR	20-SEP-1999;	99US-0154779P.	Qy	1	AEEGAVACVVRPLN-	-----SREES-----	-----LGETA-----QVYWKTDN
PR	02-JUL-1999;	99US-0142055P.	PR	22-SEP-1999;	99US-0155139P.	Db	7	SERDSISVTVRFLRPLRYARSDLDAMDQIYCRDYSFHVDAIGVNSLLGEVQRGDEVAVYPDG	-----	-----
PR	06-JUL-1999;	99US-0142390P.	PR	23-SEP-1999;	99US-0155486P.	Qy	35	NVI--YQVGDGSKSFNDRVFHGNBETTKNVYEEIAAPIDSAIQYNGTIFAYGOTASGKT	-----	-----
PR	08-JUL-1999;	99US-0142803P.	PR	24-SEP-1999;	99US-0155659P.	Db	67	DTLVRHEYNPLTAVAFDKVFGPQATTIDVYDVARPVVKAMEGVNGTVFAYGVTSSGKT	-----	-----
PR	09-JUL-1999;	99US-0142920P.	PR	28-SEP-1999;	99US-0156458P.	Qy	93	YTM-----	-----	-----MGSEDLGVIPRAIHDF
PR	12-JUL-1999;	99US-0142977P.	PR	29-SEP-1999;	99US-0156596P.	Db	127	HTMVRVLKKNPYTPFSLVFLQNTCANWDLFFNLVPLWLLGDQSPGIIPLAIXDVF	-----	-----
PR	13-JUL-1999;	99US-0143542P.	PR	04-OCT-1999;	99US-0157117P.	Qy	114	QKIKKF-----	-----	-----PDREFLLRVSYMEIYNETITDLCGTQKMKPLIIRDNVNRNVVADLT
PR	14-JUL-1999;	99US-0143624P.	PR	05-OCT-1999;	99US-0157753P.	Db	187	SIIOQVSLNGTGPREFLLRVSYLEIYNEVINDLLDPTG--QNLVRRED-SQGYVEGK	-----	-----
PR	15-JUL-1999;	99US-0144005P.	PR	06-OCT-1999;	99US-0157865P.	Qy	168	EEVYVTSMAKWKTKGSKSHYGETKNQORSSSHITFRMILESEKGEPSNCEGSKV	-----	-----
PR	16-JUL-1999;	99US-0144085P.	PR	07-OCT-1999;	99US-0158023P.	Db	244	EEVVLSPGHLSFIAAGEBHRHVSNNFNLLSSRSHITFTLMVESATGDEYD---GVIF	-----	-----
PR	19-JUL-1999;	99US-0144332P.	PR	12-OCT-1999;	99US-0158232P.	Qy	228	SHLNLVDLAGERRAQTCGAAGVRLKEGNCINRSLFIPILGOVIKKLSDGOVGGGFINYRDSKL	-----	-----
PR	19-JUL-1999;	99US-0144333P.	PR	13-OCT-1999;	99US-0159294P.	Db	301	SQNLNLDLAGSE-SSKTETTGLRRKEGSIYINKSLTLTGLTICKLSEK-ATHIPYRDSKL	-----	-----
PR	19-JUL-1999;	99US-0144334P.	PR	13-OCT-1999;	99US-0159295P.					
PR	19-JUL-1999;	99US-0144335P.	PR	14-OCT-1999;	99US-0159330P.					
PR	20-JUL-1999;	99US-0144352P.	PR	14-OCT-1999;	99US-0159331P.					
PR	20-JUL-1999;	99US-0144632P.	PR	14-OCT-1999;	99US-0159637P.					
PR	20-JUL-1999;	99US-0144844P.	PR	18-OCT-1999;	99US-0159584P.					
PR	21-JUL-1999;	99US-0145086P.	PR	21-OCT-1999;	99US-0160741P.					
PR	21-JUL-1999;	99US-0145088P.	PR	21-OCT-1999;	99US-0160767P.					
PR	22-JUL-1999;	99US-0145087P.	PR	21-OCT-1999;	99US-0160770P.					
PR	22-JUL-1999;	99US-0145089P.	PR	21-OCT-1999;	99US-0160814P.					
PR	22-JUL-1999;	99US-0145192P.	PR	21-OCT-1999;	99US-0160815P.					
PR	23-JUL-1999;	99US-0145145P.	PR	22-OCT-1999;	99US-0160980P.					
PR	23-JUL-1999;	99US-0145218P.	PR	22-OCT-1999;	99US-0160981P.					
PR	26-JUL-1999;	99US-0145276P.	PR	25-OCT-1999;	99US-0161404P.					
PR	27-JUL-1999;	99US-0145913P.	PR	25-OCT-1999;	99US-0161405P.					
PR	27-JUL-1999;	99US-0145918P.	PR	25-OCT-1999;	99US-0161406P.					
PR	27-JUL-1999;	99US-0145919P.	PR	26-OCT-1999;	99US-0161359P.					
PR	28-JUL-1999;	99US-0145951P.	PR	26-OCT-1999;	99US-0161360P.					
PR	02-AUG-1999;	99US-0146386P.	PR	26-OCT-1999;	99US-0161361P.					
PR	02-AUG-1999;	99US-0146388P.	PR	28-OCT-1999;	99US-0161920P.					
PR	03-AUG-1999;	99US-0147038P.	PR	28-OCT-1999;	99US-0161922P.					
PR	04-AUG-1999;	99US-0147204P.	PR	28-OCT-1999;	99US-0161933P.					
PR	04-AUG-1999;	99US-0147302P.	PR	29-OCT-1999;	99US-0162142P.					
PR	05-AUG-1999;	99US-0147192P.								
PR	05-AUG-1999;	99US-0147260P.								
PR	06-AUG-1999;	99US-0147303P.								
PR	06-AUG-1999;	99US-0147416P.								
PR	09-AUG-1999;	99US-0147493P.								
PR	09-AUG-1999;	99US-0147935P.								
PR	10-AUG-1999;	99US-0148171P.								
PR	11-AUG-1999;	99US-0148319P.								
PR	12-AUG-1999;	99US-0148341P.								
PR	13-AUG-1999;	99US-0148565P.								
PR	13-AUG-1999;	99US-0148684P.								
PR	16-AUG-1999;	99US-0149368P.								
PR	17-AUG-1999;	99US-0149175P.								
PR	18-AUG-1999;	99US-0149426P.								
PR	20-AUG-1999;	99US-0149722P.								
PR	20-AUG-1999;	99US-0149723P.								
PR	20-AUG-1999;	99US-0149929P.								
PR	23-AUG-1999;	99US-0149902P.								
PR	23-AUG-1999;	99US-0149930P.								
PR	25-AUG-1999;	99US-0150566P.								
PR	26-AUG-1999;	99US-0150884P.								
PR	27-AUG-1999;	99US-0151065P.								
PR	27-AUG-1999;	99US-0151066P.								
PR	27-AUG-1999;	99US-0151080P.								
PR	30-AUG-1999;	99US-0151303P.								
PR	31-AUG-1999;	99US-0151438P.								
PR	01-SEP-1999;	99US-0151930P.								
PR	07-SEP-1999;	99US-0152363P.								

```
QY 288 TRILONSLGNPKTRIITCTPV--SPDETITALQFASAKYKMTQPYVNEVSTDEALK 345
Db 359 TRLOSLSGHGHVSLCTITTPASSSEETHNTLKFAFRAKSIEIVASRNQIIDEKSLIK 418
QY 346 RYRKEIMDLKKQLEEV-----SLETRAQAMEKQ--LAQLLEEKDLLOKQVNEK 392
Db 419 KYQREISTLKLDELQRLRGMLGVGSHEELMSLKQOLEEGOVQMSRLEEEERAKAALMSR 478
QY 393 IENITRLMLVTSS 404
Db 479 IQKTLKLILVST 490

RESULT 14
AAC31111
ID AAG31111 standard; protein; 1069 AA.
XX AC AAG31111;
XX DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 37308.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-00301439.
XX 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 04-MAY-1999; 99US-0132407P.
PR 05-MAY-1999; 99US-0132484P.
PR 06-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 11-MAY-1999; 99US-0132863P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 18-MAY-1999; 99US-0134370P.
PR 19-MAY-1999; 99US-0134768P.
PR 20-MAY-1999; 99US-0134941P.
PR 21-MAY-1999; 99US-0135124P.
PR 24-MAY-1999; 99US-0135353P.
PR 25-MAY-1999; 99US-0135629P.
PR 27-MAY-1999; 99US-0136021P.
PR 28-MAY-1999; 99US-0136392P.
PR 01-JUN-1999; 99US-0136782P.
PR 03-JUN-1999; 99US-0137222P.
PR 04-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 14-JUN-1999; 99US-0138847P.
PR 16-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 17-JUN-1999; 99US-0139453P.
PR 18-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 21-JUN-1999; 99US-0139763P.
PR 22-JUN-1999; 99US-0139817P.
PR 23-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 24-JUN-1999; 99US-0140354P.
PR 28-JUN-1999; 99US-0140695P.
PR 29-JUN-1999; 99US-0140823P.
PR 30-JUN-1999; 99US-0140991P.
PR 01-JUL-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 02-JUL-1999; 99US-0142154P.
PR 06-JUL-1999; 99US-0142055P.
PR 08-JUL-1999; 99US-0142390P.
PR 09-JUL-1999; 99US-0142803P.
PR 12-JUL-1999; 99US-0142920P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 20-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144632P.
PR 21-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144881P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 26-JUL-1999; 99US-0145224P.
PR 27-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 28-JUL-1999; 99US-0145919P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 05-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
```


Fri Aug 6 10:49:09 2004

```
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 21-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 33.8%; Score 694; DB 3; Length 1121;
Best Local Similarity 36.2%; Pred. No. 7.2e-52;
Matches 178; Conservative 87; Mismatches 131; Indels 96; Gaps 14;

QY 1 AEGAVAVCVVRPLN-----SREES-----LGETA---OVYWKTDN 34
Db 94 SERDSISVTVRFRPLRYARSDLAMDOIYCRDYSFHVDAIGVNSLLGEYQRGDEVANYPDG 153
QY 35 NVI--YQVDSKSFNDRVPHGNETTKNVYEEIAAPIIDSAIOQYNGTTFAYGQTASGKT 92
Db 154 DTLVRHEYNPLTAYAFDKVFPQATTIDVDVAARPVVKAAMEGVNGTVFAYGVTSSGKT 213
QY 93 YTM-----MGSEDLGVIPRAHDI 113
Db 214 HTMHVRLVKKKPYTPPSLVFLQNTCAMNDLFFNLVPLWYLLGDQSPGIIPLAIKOVF 273
QY 114 QTKKF-----PDREFLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNVYVADLT 167
Db 274 SIIQVDSGLNGTGPREFLLRVSYLEIYNEVINDLLDPTG--QNLVRRED--SQGYVEGK 330
QY 168 BEVYVTSEMALKMTTKGKSRHYGETKMNORSRSHITFRMILESRKGPSPNCEGSKV 227
Db 331 BEVVLSPCHALSFAAGEHRHVGSNNFNLLSRSHITFTLMVRESSATGDEYD--GVIF 387
QY 228 SHLNVDLAGSRAAQTGAAGVRLKEGCNINRSFILGOVTKLSDQGVGFNYRDSKL 287
Db 388 SOLNIDLAGE--SSKTETTLRRKKGYSYNKSLTLTGTVIGKLGK--ATHIPYRDSKL 445
QY 288 TRILQNSLGNPKTRITCTTPV--SPDETTLAQEASTAKYMKNTPYNEVSTDEALLK 345
Db 446 TRILQSSLSGHGHSVSLICTITTPASSSEETHTLKFPASRAKSIETIYASRNOIIEKSLIK 505
QY 346 RYRKIMDLKOLEEV-----SLETRAQAMEKDQ--LAQLLEEKDLLQKVQNEK 392
Db 506 KYQREISTKLKQLQRLRGMLVGVSHLEMLSLKQOLEEGQVQMQRSLREEEEAALMSR 565
QY 393 IENLRLMLVTSS 404
Db 566 IQKLTKLILVST 577
```

Search completed: July 29, 2004, 09:35:13
Job time : 25.3176 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 09:30:02 ; Search time 7.15731 Seconds
(without alignments)
2914.068 Million cell updates/sec

Title: US-10-045-631b-88_COPY_2_405

Perfect score: 2053

Sequence: 1 AEEGAVACVVRPLNSREE.....LQKVQNEKIENLRMLVTSS 404

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A COMB.pap:*
2: /cgn2_6/ptodata/2/iaa/5B COMB.pap:*
3: /cgn2_6/ptodata/2/iaa/6A COMB.pap:*
4: /cgn2_6/ptodata/2/iaa/6B COMB.pap:*
5: /cgn2_6/ptodata/2/iaa/PCITUS COMB.pap:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2053	100.0	2662	4	US-09-595-684B-31
2	1468.5	71.5	2954	4	US-09-150-867-1
3	694.5	33.8	473	4	US-09-592-054-6
4	694.5	33.8	522	4	US-09-592-054-4
5	684.5	33.3	1231	4	US-09-595-684B-23
6	682.5	33.2	1234	4	US-09-592-054-8
7	678.5	33.0	1232	4	US-09-592-054-2
8	662.5	32.3	1066	3	US-09-541-782-8
9	662.5	32.3	1066	4	US-09-723-820-8
10	662.5	32.3	1066	4	US-10-270-085-8
11	652	31.8	411	2	US-08-713-815A-4
12	652	31.8	441	2	US-08-713-815A-3
13	652	31.8	975	4	US-09-914-259-19
14	652	31.8	1388	4	US-09-572-191-2
15	652	31.8	1388	4	US-09-723-262-2
16	652	31.8	1388	4	US-09-723-219-2
17	645	31.4	1031	4	US-09-914-259-24
18	637	31.0	967	4	US-09-914-259-21
19	634.5	30.9	1032	4	US-09-914-259-26
20	632	30.8	409	4	US-09-572-191-6
21	632	30.8	409	4	US-09-572-191-2
22	632	30.8	409	4	US-09-723-219-6
23	631.5	30.8	957	4	US-09-914-259-16
24	630.5	30.7	956	4	US-09-914-259-17
25	630.5	30.7	963	4	US-09-914-259-22
26	630.5	30.7	963	4	US-09-914-259-22
27	629.5	30.7	963	4	US-09-914-259-20

Sequence 18, Appl
Sequence 23, Appl
Sequence 29, Appl
Sequence 10, Appl
Sequence 10, Appl
Sequence 2, Appl
Sequence 6, Appl
Sequence 6, Appl
Sequence 8, Appl
Sequence 8, Appl
Sequence 4, Appl
Sequence 4, Appl
Sequence 25, Appl
Sequence 2, Appl

US-09-914-259-18
US-09-914-259-23
US-09-595-684B-29
US-09-541-782-10
US-09-723-820-10
US-10-270-085-10
US-09-428-156B-2
US-09-724-519-6
US-09-592-037-6
US-09-428-156B-6
US-09-724-519-8
US-09-592-037-8
US-09-428-156B-8
US-09-572-191-4
US-09-723-262-4
US-09-723-219-4
US-09-914-259-25
US-09-724-517-2

ALIGNMENTS

RESULT 1

US-09-595-684B-31
; Sequence 31, Application US/09595684B
; Patent No. 6544766

GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe

; APPLICANT: Ohashi, Cara

; APPLICANT: Sakowicz, Roman

; APPLICANT: Vaisberg, Eugeni

; APPLICANT: Wood, Kenneth

; APPLICANT: Yu, Ming

; TITLE OF INVENTION: Human kinesins and methods of producing
; FILE REFERENCE: cytop036

; CURRENT APPLICATION NUMBER: US/09/595,684B

; CURRENT FILING DATE: 2002-06-24

; PRIOR APPLICATION NUMBER: 09/295,612

; PRIOR FILING DATE: 2000-04-20

; NUMBER OF SEQ ID NOS: 105

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 31

; LENGTH: 2662

; TYPE: PRT

; ORGANISM: Human

US-09-595-684B-31

Query Match 100.0%; Score 2053; DB 4; Length 2662;

Best Local Similarity 100.0%; Pred. No. 1.9e-182;

Matches 404; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	AEEGAVACVVRPLNSREE	SLGETAQVYWKTNVLYQVDSKSFNDRVFGHNETTKN	60
Db	2	AEEGAVACVVRPLNSREE	SLGETAQVYWKTNVLYQVDSKSFNDRVFGHNETTKN	61
Qy	61	VYEIAAPIIDSAIQGYNGTIFAYGQTASGKTYTMGSEDLGVIPRAIHDIQKIKKFP	120	
Db	62	VYEIAAPIIDSAIQGYNGTIFAYGQTASGKTYTMGSEDLGVIPRAIHDIQKIKKFP	121	
Qy	121	DREFLRVSYMEYINETITDLCGTQMKPLIREDVNRNVYVADLITEEVVYTSMAKWK	180	
Db	122	DREFLRVSYMEYINETITDLCGTQMKPLIREDVNRNVYVADLITEEVVYTSMAKWK	181	
Qy	181	ITKGEKSRHYGETKMNORSRSHITFRMTLESREKEPCNCEGSKVSHNLVLDLAGSER	240	
Db	182	ITKGEKSRHYGETKMNORSRSHITFRMTLESREKEPCNCEGSKVSHNLVLDLAGSER	241	
Qy	241	AAQTGAAGVRLKEGNCINRSLFILQVYKLSGQVGGFINTYRDSKLTTRILQNSLGGNPK	300	
Db	242	AAQTGAAGVRLKEGNCINRSLFILQVYKLSGQVGGFINTYRDSKLTTRILQNSLGGNPK	301	

Fri Aug 6 10:49:09 2004

QY 301 TRIICITTPVDFETLTALQFASAKYMKNTPYVNEVSTDEALLKRYKEIMDLKKQLEE 360
 Db 302 TRIICITTPVDFETLTALQFASAKYMKNTPYVNEVSTDEALLKRYKEIMDLKKQLEE 361
 QY 361 VSLFRAQAMEKDQALAEKDLQKQVNEKIENLTRLMLVTSS 404
 Db 362 VSLFRAQAMEKDQALAEKDLQKQVNEKIENLTRLMLVTSS 405

RESULT 2
 US-09-150-867-1
 ; Sequence 1, Application US/09150867
 ; Patent No. 6645748
 ; GENERAL INFORMATION:
 ; APPLICANT: Wood, Kenneth W.
 ; APPLICANT: Sakowicz, Roman
 ; APPLICANT: Goldstein, Lawrence S.B.
 ; APPLICANT: Cleveland, Don W.
 ; APPLICANT: The Regents of the University of California
 ; TITLE OF INVENTION: Plus End-directed Microtubule Motor Required for
 ; FILE REFERENCE: 18557C-000110US
 ; CURRENT APPLICATION NUMBER: US/09/150,867
 ; CURRENT FILING DATE: 1998-09-10
 ; EARLIER APPLICATION NUMBER: US 60/058,645
 ; EARLIER FILING DATE: 1997-09-11
 ; NUMBER OF SEQ ID NOS: 11
 ; SOFTWARE: Patent In Ver. 2.0
 ; SEQ ID NO 1
 ; LENGTH: 2954
 ; TYPE: PRT
 ; ORGANISM: Xenopus sp.
 ; FEATURE:
 ; OTHER INFORMATION: Xenopus centromere-associated protein-E (XCENP-E)
 ; OTHER INFORMATION: member of the kinesin superfamily of microtubule
 ; OTHER INFORMATION: motor proteins
 ; FEATURE:
 ; NAME/KEY: DOMAIN
 ; LOCATION: (1)..(472)
 ; OTHER INFORMATION: kinesin like motor domain
 ; FEATURE:
 ; NAME/KEY: DOMAIN
 ; LOCATION: (473)..(2752)
 ; OTHER INFORMATION: rod domain
 ; FEATURE:
 ; NAME/KEY: DOMAIN
 ; LOCATION: (2753)..(2954)
 ; OTHER INFORMATION: tail domain
 ; US-09-150-867-1

Query Match 71.5%; Score 1468.5; DB 4; Length 2954;
 Best Local Similarity 70.9%; Pred. No. 1.1e-127; Indels 7; Gaps 3;
 Matches 290; Conservative 52; Mismatches 60;
 QY 1 ABEGAVCVVRPLNSREESLG-ETAQVYWKTDNNVYQVDSKGFNEDRVPHGHETHKN 60
 Db 2 SEGDAVVCVVRPLNRIQEQ--GDQANLQKAGNNTISQVDTGKGFNEDRVFNSHESQ 59
 QY 61 VVEEIAAPLIDSAIQYNGTIFAYGATAGKTYTMGSDHGLVTPRAITHDIFQKIKFP 120
 Db 60 IYQEIAPVIRGALQYNGTIFAYGATAGKTYTMGSDHGLVTPRAITHDIFQKIKFP 119
 QY 121 DREFLLRVSMYMEIYNETIIDLCTGQKMKPLIREDVNRNVYVADLTVEEVYVTSMAK 180
 Db 120 NREFLLRVSMYMEIYNETIIDLCTGQKMKPLIREDVNRNVYVADLTVEEVYVTSMAK 179
 QY 181 ITKGEKSRHYGETKMKQSRSHSTIFRMILESRKEGEPSS---NCEGSKVSHNLVDLAG 237
 Db 180 IKKGEKSRHYGETKMKQSRSHSTIFRMILESRKEGEPSS---NCEGSKVSHNLVDLAG 239
 QY 238 SERAAQTGAAGVRLKEGNCNINRSLFILGVIKKLSGQVGFNRYRDSKLTILQNSLGG 297
 Db 240 SERASQTGAAGVRLKEGNCNINRSLFILGVIKKLSGQVGFNRYRDSKLTILQNSLGG 299

QY 298 NPKTRIICTTPVDFETLTALQFASAKYMKNTPYVNEVSTDEALLKRYKEIMDLKKQ 357
 Db 300 NAKTVIICITTPVDFETLTALQFASAKYMKNTPYVNEVSTDEALLKRYKEIMDLKKQ 359
 QY 358 LE--EVSLETRAQAMEKDQALAEKDLQKQVNEKIENLTRLMLVTSS 404
 Db 360 LENLESSETRAQAMEKAEKHTQLLAIEIKQLHKEKREDRIWHLTNIVVASS 408

RESULT 3
 US-09-592-054-6
 ; Sequence 6, Application US/09592054
 ; Patent No. 6440684
 ; GENERAL INFORMATION:
 ; APPLICANT: Beraud, Christophe
 ; APPLICANT: Finer, Jeffrey
 ; APPLICANT: Sakowicz, Roman
 ; APPLICANT: Wood, Kenneth
 ; TITLE OF INVENTION: No. 6440684el motor proteins and methods for
 ; TITLE OF INVENTION: their use
 ; FILE REFERENCE: 1016
 ; CURRENT APPLICATION NUMBER: US/09/592,054
 ; CURRENT FILING DATE: 2000-07-20
 ; NUMBER OF SEQ ID NOS: 8
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 6
 ; LENGTH: 473
 ; TYPE: PRT
 ; ORGANISM: Human
 ; US-09-592-054-6

Query Match 33.8%; Score 694.5; DB 4; Length 473;
 Best Local Similarity 39.6%; Pred. No. 1.7e-56;
 Matches 169; Conservative 66; Mismatches 145; Indels 45; Gaps 8;
 QY 6 VAVCVVRPLNSREESLG-ETAQVYWKTDNNVYQVDSKGFNEDRVPHGHETHKN 64
 Db 8 VRVALRCPLVPKEISEGCMCLSFVPGEPQVVGTD--KSFTYDFVDFPSTQREVENT 65
 QY 65 IAPPIIDSAIQYNGTIFAYGATAGKTYTMG-----SEDHLGVIPRAIHDFQKIK 117
 Db 66 AVAPLIKGVFKGNATVLAQGTCGSKTYSMGGAATAEQENEPVGVIPVLIQLLFEID 125
 QY 118 KPPDREFLLRVSMYMEIYNETIIDLCTGQKMKPLIREDVNRNVYVADLTVEEVYVTSMA 177
 Db 126 KKSDFEFTLVKVSYLEIYNEEILDLCPREKQINREDPKGKIVGLTEKTVLVALDT 185
 QY 178 LKWTIKGEKSRHYGETKMKQSRSHSTIFRMILESRKEGEPSSNCEGSKVSHNLVDLAG 237
 Db 186 VSCLEQGNNSRTVASTAMNSQSRSHAFTTISLEQKESD---KNSSFRSKLHVLVDLAG 241
 QY 238 SERAAQTGAAGVRLKEGNCNINRSLFILGVIKKLSGQVGFNRYRDSKLTILQNSLGG 297
 Db 242 SERQKTKAEGDRLEKGINNGLICLGNVISALGDDKGGFVPEYDSKLTILQNSLGG 301
 QY 298 NPKTRIICTITPV--SFDETLTALQFASAKYMKNTPYVNEVSTDEALLKRYKEIMDLK 355
 Db 302 NSHTLMIACVSPADSNLEETLTAVADRARKIKKPIVN-----IDPQTAENHLK 353
 QY 356 KQLEEVSL-----ETRAQAMEKDQALAEKDLQKQVNEKIENL 396
 Db 354 QVQVQLVLLQAHGGLPGSITVPFENLQSLMEKQ--SLVEENEKLSRGLSEAGQT 411
 QY 397 TRML 400
 Db 412 AQML 415

RESULT 4
 US-09-592-054-4
 ; Sequence 4, Application US/09592054
 ; Patent No. 6440684


```

; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Finer, Jeffrey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Wood, Kenneth
; TITLE OF INVENTION: No. 6440684el motor proteins and methods for
; FILE REFERENCE: 1016
; CURRENT APPLICATION NUMBER: US/09/592,054
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 522
; TYPE: PRT
; ORGANISM: Human
US-09-592-054-4

Query Match      33.8%; Score 694.5; DB 4; Length 522;
Best Local Similarity 39.6%; Pred. No. 2e-56;
Matches 168; Conservative 66; Mismatches 145; Indels 45; Gaps 8;

QY 6 VAVCVVRPLNSREESLG-ETAQVYWKTDNNVIYQVDSKSFNFDVFGHNETTKNVEE 64
Db 30 VRVALRCRLVPKEISEGCMCLSFVPGEPQVVVGTD--KSFTYDFVDFPSTEQEEVENT 87
QY 65 IAAPIDSALIOGYNGTIFAYGQTASGKTYTMG-----SEDLGVIPRAIHDFQKIK 117
Db 88 AVAPLIKGVFKGNATVLAYGQTGSKTYSMGAYTAEOENPTVGVIPRVQLLFKEID 147
QY 118 KPDPREFLLRVSYMEIYNETITDLICGQKMKPLIREDVNRNVVADLTERVVTSEMA 177
Db 148 KKSDFEFTLKVSYLEIYNEEILDLLCPREKAQINIREDPKEGKIVGLTEKTVLVALDT 207
QY 178 LKWTIKGKSRHYGKTKNQRSRSHITFRMILESRKEGPNCSGVKVSHLNVLVDLAG 237
Db 208 VSCLEQGNNSRTVASTAMNSQSSRSHAFTISLEQRKSD-----KNSFRSKLHLVDLAG 263
QY 238 SERAAQGAAGVRLKEGNCINRSLFILGVIKLSDGVGGFINVRDLSKLTILQNSLGG 297
Db 264 SERQKTKAEGDRLKEGINRGLCLGNVISALGDDKGGFAPYRDSKLTLLQDSLGG 323
QY 298 NPKTRIICITIPV--SPDETITAIQFASTAKYMKNTPYVNEVSTDEALLKRYKEIMDLK 355
Db 324 NSHTLMITACVSPADSNLEETLTLRYADRARKIKNKPVN-----IDPQTAEHLX 375
QY 356 KQLEEVSL-----ETRAQAMEKQOLAOLLEKOLLQKQVQNEKIENL 396
Db 376 QVQVQLQVLLIQAAGGTLPGSITVEPSENLSQSLMEKNQ--SLVBENEKLSRGLSEAAAGT 433
QY 397 TRML 400
Db 434 AQML 437

RESULT 5
US-09-595-684B-23
; Sequence 23, Application US/09595684B
; Patent No. 6544766
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Ohashi, Cara
; APPLICANT: Sakowicz, Roman
; APPLICANT: Vaisberg, Eugeni
; APPLICANT: Wood, Kenneth
; APPLICANT: Yu, Ming
; TITLE OF INVENTION: Human kinesins and methods of producing
; FILE REFERENCE: Cyt0036
; CURRENT APPLICATION NUMBER: US/09/595,684B
; CURRENT FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: 09/295,612
; PRIOR FILING DATE: 2000-04-20

; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Finer, Jeffrey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Wood, Kenneth
; TITLE OF INVENTION: No. 6440684el motor proteins and methods for
; FILE REFERENCE: 1016
; CURRENT APPLICATION NUMBER: US/09/592,054
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1234
; TYPE: PRT
; ORGANISM: Human
US-09-592-054-8

Query Match      33.2%; Score 682.5; DB 4; Length 1234;
Best Local Similarity 38.8%; Pred. No. 9.8e-55;
Matches 166; Conservative 73; Mismatches 142; Indels 49; Gaps 10;

QY 6 VAVCVVRPLNSREESLG-ETAQVYWKTDNNVIYQVDSKSFNFDVFGHNETTKNVEE 64

```

238 SERAQTGAAGVRLKGCNINRSIFLQGVIKLSDGVGGVFNTRDLSKLTILQNSLGG 297
 244 SERQKTKAEGDRLKGCNINRGLLGNVISALGDDKGGFAPYRDSKLTLLQDSLGG 303
 298 NPKTRIICTITPV--SFDETLTALQFASTAKYMKNTPYVNEVSTDEALLKRYRKEIMDLK 355
 304 NSHTLMIACVSPADSNLEETLTLRYADRARKIKNKPVN-----IDPOTAEHLNHLK 355
 356 KOLEEVS-----ETRAQAMEKDQALAEKDLQKQVNEKIENL 396
 356 QOVQQLQVLLQAHGGLPGSITVPESENLOSIMKQ--SLVEENEKLSRGLSEAGQT 413
 397 TRLM 400
 414 AQML 417

RESULT 8
 US-09-541-782-8
 ; Sequence 8, Application US/09541782
 ; Patent No. 6284480
 ; GENERAL INFORMATION:
 ; APPLICANT: Nislow, Corey
 ; APPLICANT: Sakowicz, Roman
 ; APPLICANT: Beraud, Christophe
 ; TITLE OF INVENTION: Antifungal assay
 ; FILE REFERENCE: 1015
 ; CURRENT APPLICATION NUMBER: US/09/541,782
 ; CURRENT FILING DATE: 2000-04-03
 ; NUMBER OF SEQ ID NOS: 10
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 8
 ; LENGTH: 1066
 ; TYPE: PRT
 ; ORGANISM: Drosophila melanogaster
 US-09-541-782-8

Query Match 32.3%; Score 662.5; DB 3; Length 1066;
 Best Local Similarity 37.4%; Pred. No. 5.8e-53;
 Matches 159; Conservative 78; Mismatches 141; Indels 47; Gaps 11;

6 VAVCVVRPLNSREESL--GETAQVYWKTDNNVIYQVDG--SKSFNDRVPHGNETTKNV 61
 20 IQVTVVRPLNSRERCIRSAEVDVVGPREVWTRHTLDSKLTKKFTFDRSFGPSKQCDV 79
 62 YEEIAPIIDSAIOGYNGTIFAYGOTASGKTYTMMGSE-----DHLGVIPRAIH 110
 80 YSVVVSPLIEVLNGYNTVFAYGOTGKTHTWGNETAEKLSKSWEDSDIGIIPRLS 139
 111 DIFQKIKFPDREFLLRVSYMEIYNETITDILCGTQRMKPLIREDVNR--VYVADLTEE 169
 140 HLFDELMM--EYETMRISYILEYNEELCDLLSTDDTTKIRIFDDSTKKGSVIIQGLEI 198
 170 VVYVSEMAKWIITKGEKSRHYGETMQRSSRSHITFRMILESREKGEPSNCEGSKVYSH 229
 199 PVHSKDDVYKLEKGERKKTATILMAQSSRSHTVFSIVVHIRENGIEG--EDMLKIGK 256
 230 LNLVDLAGSRAAQG--AAGVRLKGCNINSLFQGVIKLSDGVGGVFNTRDLSKLT 288
 257 LNLVDLAGSENKAGNEKIRVRETVINOSLLTLGRVITLVDR--APHVPYRESKLT 314
 289 RLQNSLGNPKTRIICTITP--VSFDETLTALQFASTAKYMKNTPYVNEVSTDEALLK 346
 315 RLQNSLGRYKTSIIATISPGHKDIEETLSLEYAHRKAKNTQNKPEVNOKLTKKTVLKE 374
 347 YKEEIMDLKKQL-----EVSLETRAAMEKDQALAEKDLQKQVNEKIENL 390
 375 YTEEIDKRLDMAARDKNGIYLABETTYGEITLKLESQNR-----LNEKMLLLKALK 427
 391 EKIN 395
 428 DELQ 432

10 VRVALRCRPLVPKEISEGQCLSFVPGETQVVVGT--KSFTYDFVDPCTEQEVENK 67
 65 IAAPLIDSAIOGYNGTIFAYGOTASGKTYTMMG-----SEDHLGVIPRAIHDFQKIK 117
 68 AVAPLIKGFKNATVLAAYGOTSGKTYSMGAYTAEQENEPVGLIPRVQLLFKEID 127
 118 KFPDREFLLRVSYMEIYNETITDILCGTQRMKPLIREDVNRNVVADLTEEVVYSEMA 177
 128 QKSDFEFTLKVSYLEIYNEEILDLCPREKAQINIREDPKGIKIVGLTEKTVLVALDT 187
 178 LKWIITKGEKSRHYGETMQRSSRSHITFRMILESREKGEPSNCEGSKVSHNLVLDLAG 237
 188 VSLCEGNNRSRVASTAMNSQSSRSHAITI---SLEQKKSKDNSSFR--SKLHLVLDLAG 243
 238 SERAQTGAAGVRLKGCNINRSIFLQGVIKLSDGVGGVFNTRDLSKLTILQNSLGG 297
 244 SERQKTKAEGDRLKGCNINRGLLGNVISALGDDKGGFAPYRDSKLTLLQDSLGG 303
 298 NPKTRIICTITPV--SFDETLTALQFASTAKYMKNTPYVNEVSTDEALLKRYRKEIMDLK 355
 304 NSHTLMIACVSPADSNLEETLTLRYADRARKIKNKPVN-----IDPHTAEHLNHLK 355
 356 KOLEEVS-----ETRAQAMEKDQALAEKDLQKQVNEKIENL 396
 356 QOVQQLQVLLQAHGGLPGSITVPESENLOSIMKQ--SLVEENEKLSRGLSEAGQT 413
 393 IENLRLMVT 402
 414 AQMLERIIILT 423

RESULT 7
 US-09-592-054-2
 ; Sequence 2, Application US/09592054
 ; Patent No. 6440684
 ; GENERAL INFORMATION:
 ; APPLICANT: Beraud, Christophe
 ; APPLICANT: Finer, Jeffrey
 ; APPLICANT: Sakowicz, Roman
 ; APPLICANT: Wood, Kenneth
 ; TITLE OF INVENTION: their use
 ; FILE REFERENCE: 1016
 ; CURRENT APPLICATION NUMBER: US/09/592,054
 ; CURRENT FILING DATE: 2000-07-20
 ; NUMBER OF SEQ ID NOS: 8
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 2
 ; LENGTH: 1232
 ; TYPE: PRT
 ; ORGANISM: Human
 US-09-592-054-2

Query Match 33.0%; Score 678.5; DB 4; Length 1232;
 Best Local Similarity 39.4%; Pred. No. 2.3e-54;
 Matches 167; Conservative 66; Mismatches 146; Indels 45; Gaps 9;

6 VAVCVVRPLNSREESL--GETAQVYWKTDNNVIYQVDGSKSFNDRVPHGNETTKNVYEE 64
 10 VRVALRCRPLVPKEISEGQCLSFVPGEPQVVVGT--KSFTYDFVDPCTEQEVENF 67
 65 IAAPLIDSAIOGYNGTIFAYGOTASGKTYTMMG-----SEDHLGVIPRAIHDFQKIK 117
 68 AVAPLIKGFKNATVLAAYGOTSGKTYSMGAYTAEQENEPVGLIPRVQLLFKEID 127
 118 KFPDREFLLRVSYMEIYNETITDILCGTQRMKPLIREDVNRNVVADLTEEVVYSEMA 177
 128 KKSDFEFTLKVSYLEIYNEEILDLCPREKAQINIREDPKGIKIVGLTEKTVLVALDT 187
 178 LKWIITKGEKSRHYGETMQRSSRSHITFRMILESREKGEPSNCEGSKVSHNLVLDLAG 237
 188 VSLCEGNNRSRVASTAMNSQSSRSHAITI---SLEQKKSKDNSSFR--SKLHLVLDLAG 243

```

RESULT 9
US-09-723-820-8
; Sequence 8, Application US/09723820
; Patent No. 6468760
; GENERAL INFORMATION:
; APPLICANT: Nislow, Corey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Beraud, Christophe
; TITLE OF INVENTION: Antifungal Assay
; FILE REFERENCE: 1015
; CURRENT APPLICATION NUMBER: US/09/723,820
; CURRENT FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 09/541,782
; PRIOR FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1066
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-723-820-8

Query Match      32.3%   Score 662.5; DB 4; Length 1066;
Best Local Similarity 37.4%; Pred. No. 5.8e-53;
Matches 159; Conservative 78; Mismatches 141; Indels 47; Gaps 11;

QY 6 VAVCVVRPLNSREESL--GETAQVYWKTDNNVIYQVDG--SKSFNDRVPHGNETTKNV 61
   : ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 20 IQVVVRPLNSRERCIRSAEVDVVGPREVTRHTLDSKLTKKFTFDRSGPESKQCDV 79

QY 62 YEEIAPIIDSAIQVNGTIFAYGQTASGKTYTMGSE-----DHLGVIPRAIH 110
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 80 YSVVVSPLIEVLNGYCNCTVFAYGQTGKTHTMVGNETAELKSSWEDSDIGIIPRAL 139

QY 111 DIFQIKKPPDFRFLRVSMYIYNETITDILCGTQKMKPLIREDVNR--NVYVADLTEE 169
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 140 HLFDELRLMM--EVEYTNRIISYLEYNEELCDLSTDITTKIRIFDDSTKKGSVIIQGLEEI 198

QY 170 VVYTSMALKWKITKGEKSHYGETKNQSRSSSHITFRMILESREKGEPCNCEGSKVSH 229
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 199 PVHSKDDVYKLEKGERKKTATLMAQSSSHVFSIVVHIRENGIEG--EDMLKIGK 256

QY 230 LNLVDLAGSRAAQTG--AAGVRLKEGCNINRSLFILGOVVKKLSGQGVGFYINRDSKLT 288
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 257 LNLVDLAGSENVSKAGNEKGIKRVETVNIQSLLTLGRVITLVDR--APHVPYRESKLT 314

QY 289 RILQNSLGNPKTRIICTITP--VSFDETLTALQFASAKYMKNTPYNVNEVSTDEALLKR 346
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 315 RLQESLGGRTKTSIIATISPGHKDIEETLSLEYAHRAKNIQNKPEVNQKLTKKTVLKE 374

QY 347 YRKEIMDLKKQI-----EEVSLQTRAQAMEKDQLAQLLEEKDLLOKQVN 390
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 375 YTEEDIKLRDLMAARDKNGIYLAETGYEITLKLESQNR-----LNEKMLLKALK 427

QY 391 EKIEEN 395
   : : : : :
Db 428 DELQN 432

RESULT 10
US-10-270-085-8
; Sequence 8, Application US/10270085
; Patent No. 6627408
; GENERAL INFORMATION:
; APPLICANT: Nislow, Corey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Beraud, Christophe
; TITLE OF INVENTION: Antifungal Assay
; FILE REFERENCE: 1015
; CURRENT APPLICATION NUMBER: US/10/270,085
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: US/09/723,820

```

```

; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 09/541,782
; PRIOR FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1066
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-270-085-8

Query Match      32.3%   Score 662.5; DB 4; Length 1066;
Best Local Similarity 37.4%; Pred. No. 5.8e-53;
Matches 159; Conservative 78; Mismatches 141; Indels 47; Gaps 11;

QY 6 VAVCVVRPLNSREESL--GETAQVYWKTDNNVIYQVDG--SKSFNDRVPHGNETTKNV 61
   : ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 20 IQVVVRPLNSRERCIRSAEVDVVGPREVTRHTLDSKLTKKFTFDRSGPESKQCDV 79

QY 62 YEEIAPIIDSAIQVNGTIFAYGQTASGKTYTMGSE-----DHLGVIPRAIH 110
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 80 YSVVVSPLIEVLNGYCNCTVFAYGQTGKTHTMVGNETAELKSSWEDSDIGIIPRAL 139

QY 111 DIFQIKKPPDFRFLRVSMYIYNETITDILCGTQKMKPLIREDVNR--NVYVADLTEE 169
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 140 HLFDELRLMM--EVEYTNRIISYLEYNEELCDLSTDITTKIRIFDDSTKKGSVIIQGLEEI 198

QY 170 VVYTSMALKWKITKGEKSHYGETKNQSRSSSHITFRMILESREKGEPCNCEGSKVSH 229
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 199 PVHSKDDVYKLEKGERKKTATLMAQSSSHVFSIVVHIRENGIEG--EDMLKIGK 256

QY 230 LNLVDLAGSRAAQTG--AAGVRLKEGCNINRSLFILGOVVKKLSGQGVGFYINRDSKLT 288
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 257 LNLVDLAGSENVSKAGNEKGIKRVETVNIQSLLTLGRVITLVDR--APHVPYRESKLT 314

QY 289 RILQNSLGNPKTRIICTITP--VSFDETLTALQFASAKYMKNTPYNVNEVSTDEALLKR 346
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 315 RLQESLGGRTKTSIIATISPGHKDIEETLSLEYAHRAKNIQNKPEVNQKLTKKTVLKE 374

QY 347 YRKEIMDLKKQI-----EEVSLQTRAQAMEKDQLAQLLEEKDLLOKQVN 390
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 375 YTEEDIKLRDLMAARDKNGIYLAETGYEITLKLESQNR-----LNEKMLLKALK 427

QY 391 EKIEEN 395
   : : : : :
Db 428 DELQN 432

RESULT 11
US-08-713-815A-4
; Sequence 4, Application US/08713815A
; Patent No. 5830659
; GENERAL INFORMATION:
; APPLICANT: Russell J. Stewart
; TITLE OF INVENTION: ACTIVE MICROTUBULE-BASED
; TITLE OF INVENTION: SEPARATIONS BY KINESINS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Thotipe, No. 5830659th & Western, L.I.P.
; STREET: 9035 South 700 East, Suite 200
; CITY: Sandy
; STATE: Utah
; COUNTRY: USA
; ZIP: 84070
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb Storage
; COMPUTER: AST Ascentia 900N
; OPERATING SYSTEM: DOS 6.22
; SOFTWARE: Word Perfect 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/713,815A
; FILING DATE: 13-SEP-1996
; CLASSIFICATION: 435

```

Fri Aug 6 10:49:09 2004

```
ATTORNEY/AGENT INFORMATION:
; NAME: Alan J. Howarth
; REGISTRATION NUMBER: 36,553
; REFERENCE/DOCKET NUMBER: T3214/U-2202
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (801)566-6633
; TELEFAX: (801)566-0750
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 411 amino acid residues
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-713-815A-4

Query Match 31.8%; Score 652; DB 2; Length 411;
Best Local Similarity 42.5%; Pred. No. 1.3e-52;
Matches 174; Conservative 58; Mismatches 149; Indels 28; Gaps 15;

QY 1 AEGAVAVCVVRPLNSREESLGETAQVYKTDNNV---IYQVDSKSFNDRVFGHNET 57
Db 9 AEDSIKVCV-RFRPLNDSSEKAGSKFVV--KFPNNVEENCISIAAG-KVYLFDRKVKFNAS 64
QY 58 TKNVYEIAAPIIDSAIQGYNGTIFAYGQTASGKTYTM---MGSEDLHGVIPRAIHDFQ 114
Db 65 QEKVYNEAAKSIYTDVLGYNGTIFAYGQTSSGKTHMEGVIGDSVKQGIIPRVNDIFN 124
QY 115 KIKKFP--DREFLLRVSYMEIYNETITDLLCGTQMKPLIREDVNRNVYVADLVEEVYVT 173
Db 125 HIYAMENVLEPHIKVSYEIIYMDKIRDLL-DVSKVN--LSVHEDKRVYPYVGATERFVSS 182
QY 174 SEMALKWITKGEKSRHYGETKMNORSRSHITFRMILESREKGPSCSVKVS-HLNL 232
Db 183 PEDVFEVIEEGKSNRHIATVNNHSSRSHSVFLNVKQ-----ENLENQKLSGKLYL 236
QY 233 VDLASERAAQTGAAGVRLKEGCNINRSLFGLGVKILSDGVGGFNYRDSKLTRILQ 292
Db 237 VDLAGEKVKTAGTGTVDLDEAKNINKLSALGNVISALADGN-KTHIPYRDSKLTRILQ 295
QY 293 NSLGNPKTRICTITTPVSFD--ETLTALQFASTAKYMKNTPYVNEVSTDEALLKRYRKE 350
Db 296 ESLGNARTTIVICCPASPFNESETKSLDFGRRAKTVKNVVCNEELTAEWKRYRKE 355
QY 351 ---IMDLKKQLEVSLE--TRAQMEKQOLALEEKDLIQ-KVQNEKIE 394
Db 356 KEKNARLKGKVEKLEIARWRAGETVKAEEQINMEDLMEASTPNLEVE 404

RESULT 12
US-08-713-815A-3
; Sequence 3, Application US/08713815A
; Patent No. 5830659
; GENERAL INFORMATION:
; APPLICANT: Russell J. Stewart
; TITLE OF INVENTION: ACTIVE MICROBULE-BASED
; TITLE OF INVENTION: SEPARATIONS BY KINESINS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Thorpe, No. 5830659th & Western, L.L.P.
; STREET: 9035 South 700 East, Suite 200
; CITY: Sandy
; STATE: Utah
; COUNTRY: USA
; ZIP: 84070
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
; COMPUTER: AST Ascentia 900N
; OPERATING SYSTEM: DOS 6.22
; SOFTWARE: Word Perfect 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/713,815A
; FILING DATE: 13-SEP-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:

NAME: Alan J. Howarth
REGISTRATION NUMBER: 36,553
REFERENCE/DOCKET NUMBER: T3214/U-2202
TELECOMMUNICATION INFORMATION:
TELEPHONE: (801)566-6633
TELEFAX: (801)566-0750
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 441 amino acid residues
TYPE: amino acid
TOPOLOGY: linear
US-08-713-815A-3

Query Match 31.8%; Score 652; DB 2; Length 441;
Best Local Similarity 42.5%; Pred. No. 1.5e-52;
Matches 174; Conservative 58; Mismatches 149; Indels 28; Gaps 15;

QY 1 AEGAVAVCVVRPLNSREESLGETAQVYKTDNNV---IYQVDSKSFNDRVFGHNET 57
Db 9 AEDSIKVCV-RFRPLNDSSEKAGSKFVV--KFPNNVEENCISIAAG-KVYLFDRKVKFNAS 64
QY 58 TKNVYEIAAPIIDSAIQGYNGTIFAYGQTASGKTYTM---MGSEDLHGVIPRAIHDFQ 114
Db 65 QEKVYNEAAKSIYTDVLGYNGTIFAYGQTSSGKTHMEGVIGDSVKQGIIPRVNDIFN 124
QY 115 KIKKFP--DREFLLRVSYMEIYNETITDLLCGTQMKPLIREDVNRNVYVADLVEEVYVT 173
Db 125 HIYAMENVLEPHIKVSYEIIYMDKIRDLL-DVSKVN--LSVHEDKRVYPYVGATERFVSS 182
QY 174 SEMALKWITKGEKSRHYGETKMNORSRSHITFRMILESREKGPSCSVKVS-HLNL 232
Db 183 PEDVFEVIEEGKSNRHIATVNNHSSRSHSVFLNVKQ-----ENLENQKLSGKLYL 236
QY 233 VDLASERAAQTGAAGVRLKEGCNINRSLFGLGVKILSDGVGGFNYRDSKLTRILQ 292
Db 237 VDLAGEKVKTAGTGTVDLDEAKNINKLSALGNVISALADGN-KTHIPYRDSKLTRILQ 295
QY 293 NSLGNPKTRICTITTPVSFD--ETLTALQFASTAKYMKNTPYVNEVSTDEALLKRYRKE 350
Db 296 ESLGNARTTIVICCPASPFNESETKSLDFGRRAKTVKNVVCNEELTAEWKRYRKE 355
QY 351 ---IMDLKKQLEVSLE--TRAQMEKQOLALEEKDLIQ-KVQNEKIE 394
Db 356 KEKNARLKGKVEKLEIARWRAGETVKAEEQINMEDLMEASTPNLEVE 404

RESULT 13
US-09-914-259-19
; Sequence 19, Application US/09914259
; Patent No. 6495336
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/09/914,259
; CURRENT FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 975
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; US-09-914-259-19

Query Match 31.8%; Score 652; DB 4; Length 975;
Best Local Similarity 42.5%; Pred. No. 4.9e-52;
Matches 174; Conservative 58; Mismatches 149; Indels 28; Gaps 15;

QY 1 AEGAVAVCVVRPLNSREESLGETAQVYKTDNNV---IYQVDSKSFNDRVFGHNET 57
Db 9 AEDSIKVCV-RFRPLNDSSEKAGSKFVV--KFPNNVEENCISIAAG-KVYLFDRKVKFNAS 64
```


us-10-045-631b-88_copy_2_405.ra1

Fri Aug 6 10:49:09 2004

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 09:39:47 ; Search time 20.1784 Seconds
(without alignments)
6280.361 Million cell updates/sec

Title: US-10-045-631B-88_COPY_2_405

Perfect score: 2053

Sequence: 1 AEGAVAVCVRPVPLNSREE.....LQKVQNEKIEIETRLMLVTSS 404

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/CT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/CTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	927	45.2	1382	16	US-10-437-963-176714
2	872.5	42.5	694	12	US-10-425-114-62697
3	744.5	36.3	1006	16	US-10-437-963-147933
4	718.5	35.0	420	16	US-10-437-963-114373
5	708.5	34.5	956	16	US-10-437-963-182113
6	705.5	34.4	955	12	US-10-425-114-62672
7	694.5	33.8	1232	14	US-10-116-712-670
8	694.5	33.8	1232	16	US-10-408-765A-2153
9	694.5	33.8	1235	15	US-10-334-143-8
10	691.5	33.8	1237	15	US-10-334-143-33
11	691.5	33.7	965	16	US-10-437-963-169150
12	691.5	33.7	1029	16	US-10-311-642-2
13	684.5	33.3	1232	14	US-10-116-712-664
14	684.5	33.3	1232	14	US-10-116-712-669
15	682.5	33.2	672	16	US-10-408-765A-1664

16	671.5	32.7	966	12	US-10-425-114-62697
17	663.5	32.3	406	15	US-10-369-493-3789
18	660.5	32.2	517	15	US-10-369-493-12881
19	652	31.8	975	14	US-10-080-608A-19
20	652	31.8	975	15	US-10-370-685-108
21	652	31.8	1388	12	US-10-332-089-2
22	652	31.8	1388	14	US-10-146-473-82
23	652	31.8	1388	15	US-10-173-993-32
24	652	31.8	1388	16	US-10-188-832-164
25	651.5	31.7	757	16	US-10-220-120-366
26	651.5	31.7	757	16	US-10-363-829-316
27	645	31.4	1031	14	US-10-080-608A-24
28	645	31.4	1031	15	US-10-370-685-113
29	644.5	31.4	1022	16	US-10-437-963-133108
30	644.5	31.4	1362	14	US-10-296-838-2
31	643.5	31.3	1826	16	US-10-408-765A-1477
32	643	31.3	821	16	US-10-437-963-114370
33	637	31.0	967	14	US-10-080-608A-21
34	637	31.0	967	15	US-10-370-685-110
35	634.5	30.9	1011	12	US-10-336-472-24
36	634.5	30.9	1011	12	US-10-236-417-58
37	634.5	30.9	1032	14	US-10-080-608A-26
38	634.5	30.9	1032	15	US-10-370-685-115
39	632	30.8	409	12	US-10-332-089-6
40	631.5	30.8	957	14	US-10-080-608A-16
41	631.5	30.8	957	15	US-10-370-685-105
42	631.5	30.8	1027	14	US-10-080-608A-27
43	631.5	30.8	1027	15	US-10-370-685-116
44	630.5	30.7	956	14	US-10-080-608A-17
45	630.5	30.7	956	15	US-10-370-685-106

ALIGNMENTS

RESULT 1

US-10-437-963-176714

; Sequence 176714, Application US/10437963

; Publication No. US20040123343A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.

; APPLICANT: Kovalic, David K.

; APPLICANT: Zhou, Yihua

; APPLICANT: Cao, Yongwei

; APPLICANT: Wu, Wei

; APPLICANT: Boukharov, Andrey A.

; APPLICANT: Barbazuk, Brad

; APPLICANT: Li, Ping

; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53221)B

; CURRENT APPLICATION NUMBER: US/10/437,963

; CURRENT FILING DATE: 2003-05-14

; NUMBER OF SEQ ID NOS: 204966

; SEQ ID NO 176714

; LENGTH: 1382

; TYPE: PRT

; ORGANISM: Oryza sativa

; FEATURE:

; OTHER INFORMATION: Clone ID: PAT_MRT4530_74437C.1.pep

US-10-437-963-176714

Query Match

Best Local Similarity 45.2%; Score 927; DB 16; Length 1382;

Matches 213; Conservative 62; Mismatches 108; Indels 30; Gaps 11;

QY 6 VAVCVRPVPLNSREESIGETAQYVWKTDNNVI-YQVDSKSFNDRVPHGNETTKNYEE 64

Db 4 IHVAVRARPITA--EDAGSP---WRVSGNAIALSTQPSIRFEFDRIFGECRTADVGA 58

QY 65 IAAPIDSAIQYNGTTFAYGQTASGKTYTMGSEDHGVIPRAIHDFIQKIKFPDEF 124

Db 59 RTKHIVDSVRGNGTVFAVQVNSGKTYTMRSGNEPGIIPLAVDHDFRTIEHLDREF 118

Db 327 ALGNKTSIICTAAPEEIHIEETRTLQFASRAKCVSNCAQVNEILTDAALLKQKLEI 386
 QY 352 MDLKKQ-----LEEVSLFTR-----AQAMEKDQALAEKDLQKQVNEKLENLTM 399
 Db 387 BELRLKQSHSEGLEQVVLKLRNDMHKSELERDLAMELEDERKURMTLEQHLTEQOKL 446
 QY 400 LVTSS 404
 Db 447 EAISS 451

RESULT 3

US-10-437-963-147933

; Sequence 147933, Application US/10437963

; Publication No. US20040123343A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.

; APPLICANT: Kovalic, David K.

; APPLICANT: Zhou, Yihua

; APPLICANT: Cao, Yongwei

; APPLICANT: Wu, Wei

; APPLICANT: Boukharov, Andrey A.

; APPLICANT: Barbazuk, Brad

; APPLICANT: Li, Ping

; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53221)B

; CURRENT APPLICATION NUMBER: US/10/437,963

; CURRENT FILING DATE: 2003-05-14

; NUMBER OF SEQ ID NOS: 204966

; SEQ ID NO 147933

; LENGTH: 1006

; TYPE: PRT

; ORGANISM: Oryza sativa

; FEATURE:

; NAME/KEY: unsure

; LOCATION: (1)...(1006)

; OTHER INFORMATION: unsure at all Xaa locations

; FEATURE:

; OTHER INFORMATION: Clone ID: PAT_MRT4530_48414C.1.pep

US-10-437-963-147933

Query Match 36.3%; Score 744.5; DB 16; Length 1006;

Best Local Similarity 41.7%; Pred. No. 3.7e-60;

Matches 178; Conservative 84; Mismatches 134; Indels 31; Gaps 12;

QY 1 AEGAVAVCVVRPLNREEESLGETAQVYKTDNNVYQVDSK-SFNFDVFGHNETK 59

Db 64 AAKESVAVAVRPLSPREVRGE--KIAYADGETVARSEQSNLAYAYDRVFGPTTTR 121

QY 60 NYVEIAAPIIDSALQINGTIFAYGQTASCKTYTMGSEDLGVIPRAIHDFOKIKKF 119

Db 122 HIYDAVQVYVNGAMKINGTIFAYGVTSKTHMGDQSPGVIPAVKDFINIQET 181

QY 120 PDREFLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRVYVADLTVEVYVSEMAIK 179

Db 182 PNREFLLRVSYMEIYNEVDNLL--NPAGNLRIDEIQQTI-VEGIKEEAVLSPVHALS 238

QY 180 WITKGE---KSRHYGETKONQRRSSHTIFRMILESREKPEPSCEGSKVSHNLVDLA 236

Db 239 LIAAGEVSELRHVSTNENLSSRSHTIFTLTIESPRGOSNEAE-AVTLSQLNLIDIA 297

QY 237 GSERAQTAAGVRLKEGCNINRSLFILGOVITKLSLSDGVGFIVNRDSKLTFRILQNSIG 296

Db 298 GSE-SRREVTAAGVHOGESYINKSLTLTKVLSKLT-DKATHIIFRDSKLTFLKSSLS 355

QY 297 GNPKTRITICTIPVS--FDETLTALQFASAKYKMTNPPVNEVSTDEALLKRYKEIMDL 354

Db 356 GQGRVSLICTVTPASSNSEETHNTLKFAHRAKHIEIQATQNKIMDARSLLIKKYQNEIQL 415

QY 355 KXOLEVSELETAQAMEKDQALQ--LLBEK-DLLQKQVQ-----EKLENLT 397

Db 416 KEELEQLRRSIRTPIEDTMQKHLLTETKQKSLRLEQGEAKAALLERIEHLT 475

QY 125 LLRVSYMEIYNETITDLCGTQKMKPLIREDVNRVYVADLTVEVYVSEMAIKWITKG 184

Db 119 LLRMYSMEIYNEIENDLL--VPEHRKLQIHESIERGIYVAGLRHEEIVTCEQVLEPMSFG 176

QY 185 EKSRHYGETKONQRRSSHTIFRMILESREKPEPS---NCEGSVKVSHNLVDLAGSER 240

Db 177 ESHRHIGETNMVYSSRSHTIFRMVIESREKVDSEBAGESCD-AVRVSVNLVDLAGSER 235

QY 241 AAOGTGAAGVRLKEGCNINRSLFILGOVITKLSLSDG--QVGFIVNRDSKLTFRILQNSIGN 298

Db 236 AAKTGAAGVRLKEGSHINKSLMTLGTVIKLSREGIQGGHVPYRDSKLTFRILQNALGGN 295

QY 299 PKTRIICTIT--PVSFDETLTALQFASAKYKMTNPPVNEVSTDEALLKRYKEIMDLKK 356

Db 296 ANTAIICNTITLAQVHADETKSSIQFASRALRVNACVNEILLDAALLKQKRIEELRA 355

QY 357 QL-----BEVSLTETRAQAMEKDQALAEKDLQKQVNEKIENTLRLMIVTS 403

Db 356 KLRSELEKERISLE-----LEEKKAKEQDKELIE--QAKKIENLSSILVNS 401

RESULT 2

US-10-425-114-59725

; Sequence 59725, Application US/10425114

; Publication No. US20040034888A1

; GENERAL INFORMATION:

; APPLICANT: Liu, Jingdong

; APPLICANT: Zhou, Yihua

; APPLICANT: Kovalic, David K.

; APPLICANT: Screen, Steven E

; APPLICANT: Tabaska, Jack E

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53313)B

; CURRENT APPLICATION NUMBER: US/10/425,114

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 73128

; SEQ ID NO 59725

; LENGTH: 694

; TYPE: PRT

; ORGANISM: Zea mays

; FEATURE:

; OTHER INFORMATION: Clone ID: LIB3912-011-C4_FLI.pep

US-10-425-114-59725

Query Match 42.5%; Score 872.5; DB 12; Length 694;

Best Local Similarity 46.4%; Pred. No. 2.2e-72;

Matches 197; Conservative 71; Mismatches 128; Indels 29; Gaps 9;

QY 6 VAVCVVRP-----LNSREESLGETAQVYKTDN---NVIVQVDG--SKSFNPDVRFHG 54

Db 30 ISVAVRPPNPLVAANTSPASSGGGDRWRIDDTRVSLHRAAGPITGASPAFPHVFDG 89

QY 55 NETTKVYEEAAPTIDSALQYNGTIFAYGQTASCKTYTMGSEDLGVIPRAIHDFIQ 114

Db 90 AANNRIYGTVRELVAGVGFNGTAFAYGTSKGTFTMGSDADPGIIPRAVDVFD 149

QY 115 KIKKPPDREFLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRVYVADLTVEVYVTS 174

Db 150 TVRQADREFLLRVSYMEIYNEIENDLL--TLEGQKLKHESLDRGVYVSGLREEIVNSA 207

QY 175 EMALKWIKYGEKSRHYGETKONQRRSSHTIFRMILESREKPEPSCEGSKVSHNLVD 234

Db 208 EQVFELLQGEANRHFGETNMNRSRSHITFRWVLESSEKQDQDGD-AIRVSVNLVD 266

QY 235 LAGSRAAOTGAAGVRLKEGCNINRSLFILGOVITKLSLSDG--QVGFIVNRDSKLTFRILQ 293

Db 267 LAGSERIITGAGVRLNEGKYINKSLMILGNVINKLSENGKQKHIPYRDSKLTFRILQ 326

QY 294 SLGGNPKTRITICTITP--VSFDETLTALQFASAKYKMTNPPVNEVSTDEALLKRYKEI 351

QY 398 RMLVTS 404
Db 476 ELILVA 482

RESULT 4

US-10-437-963-114373
; Sequence 114373, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 114373
; LENGTH: 420
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: unsure at all Xaa locations
; US-10-437-963-114373

Query Match 35.0%; Score 718.5; DB 16; Length 420;
Best Local Similarity 41.9%; Pred. No. 2.6e-58;
Matches 169; Conservative 72; Mismatches 129; Indels 33; Gaps 9;

QY 6 VAVCVVRPLNSREESLGETAQVYKTDNNVIYQVDS-----KSFNDFRVFHGNET 57
Db 17 IVSVRLRPVNAERAGDGD--WECAGPTTLFRGAVPERAFPASYSYDRVFSHECG 74
QY 58 TKNVVERIAPIISATQGYNGTIFAYGQTASGKTYTMGSEDHLGVIPRAIHDFQK 117
Db 75 TRQYDSGARQVAMSLGASINAFAYGQTSSTGKTYTMVGTIEY-----SMSDIYDIE 128
QY 118 KFPDREFLRYSMEIYNETITDLCGQTKMKPLIREDVNRNVVADLTEREVVTSMA 177
Db 129 KHPERFILFSAEYNEAVRDL--SSDATPLRLDDPEKGVVVEKLTETLRDKGHL 186
QY 178 LKWTKEKSRHYGETKMNQRSSRSHTIFRMILESREKGEPSNCEGSKVYVSHNLVDLAG 237
Db 187 LELAVCEAQRIGETAMNEASSRSHQILRMVTESSAKQFLGKNSSTLIACVNFVDLAG 246
QY 238 SERAAQTGAAGVRLKEGCNINRSLFILGQVTKLSDGQGVGFYINRDSKLTIRLQNSLGG 297
Db 247 SERASQTASAGRLKEGSHINRSLTLTGKVIQRLQSKGR--NGHIPYRDSKLTIRLQSSLGG 305
QY 298 NPKTRIICTTPV--SPDETALQFASAKYMKNTPYVNEVSTDEALLKRYKEIMDLK 355
Db 306 NARTALICTWSPAHCHEQSRNTLLFANCAKDVVNAQVNVVNSDQALVKHLOREIARLE 365
QY 356 KQLE---EVLETRAQAM-EKD-----QLAQLLEKDLQ 386
Db 366 NELKFPASASCTSHAELIREKDELKLNLEQLMEQKDTVQ 408

RESULT 5

US-10-437-963-182113
; Sequence 182113, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.

; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 182113
; LENGTH: 956
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(956)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_79331C.1.pap
; US-10-437-963-182113

Query Match 34.5%; Score 708.5; DB 16; Length 956;
Best Local Similarity 42.3%; Pred. No. 8e-57;
Matches 176; Conservative 73; Mismatches 138; Indels 29; Gaps 11;

QY 2 EGAVAVCVVRPLNSREESLGETAQVYKTDNN--VIYQVD-----GSKGFNDFRVFHG 54
Db 31 KEKIFVTVRVRPLSKKELAKD--QVAMECDNQITLYKGGPQDRAAPTSTYDFKVEGP 88
QY 55 NETKNVVEETAAPIISATQGYNGTIFAYGQTASGKTYTMGSEDHLGVIPRAIHDFQ 114
Db 89 ASQTEVVEEGAKOVAMSLGASINAFAYGQTSSTGKTYTMVGTIEY-----GVTESAVNDIVR 142
QY 115 KIKFPDREFLRYSMEIYNETITDLCGQTKMKPLIREDVNRNVVADLTEREVVTS 174
Db 143 HIENTPERDFIKISAMEIYNEIVKDLL--RPESTNLELDDPEKGTIVEKLEEEIAKDS 200
QY 175 EMALKWITKEKSRHYGETKMNQRSSRSHTIFRMILESREKGEPSNCEGSKVYVSHNLVD 234
Db 201 QHLRLHLSICEQVQVGETALNTSSRSHQILRMVTESSRLR-EVSGCVKSF-VANLNFVD 258
QY 235 LAGSERAQTGAAGVRLKEGCNINRSLFILGQVTKLSDGQGVGFYINRDSKLTIRLQNS 294
Db 259 LAGSERAQTGAAGVRLKEGCNINRSLTLTGKVIQRLQSKGRSGHIPYRDSKLTIRLQNS 318
QY 295 LGGNPKTRIICTTPVSVF--DETLTALQFASAKYMKNTPYVNEVSTDEALLKRYKEI 351
Db 319 LGGNARTALICTWSPAQTHVEQSRNTLFFATCAKEVTNNAKVMNVVSDQALVKHLOMEV 378
QY 352 MDLKKQLEEVSLTRAQ--AMEKDLQALLEEKDL--LQKVQN---EKIENLTRL 400
Db 379 ARLEAEALRTPDRASSSSIIIMERDRKIRQVKEKMEELKKQORDNAQSKLELQKQM 434

RESULT 6

US-10-425-114-62672
; Sequence 62672, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28

Db 244 SERQKTKAEGDRLKEGININRGILLCLGNVISALGDDKGGFVFPYRDSKLTLLQLDSLGG 303

Qy 298 NPKTRIICTTPV--SFDETLTALQFASHTAKYMKNTPYVNEVSTDEALLKRYRKEIMDLK 355

Db 304 NSHTLMACVSPADSNLEETLNTLRVADRARKINKPIVN-----IDPQTAELNHLK 355

Qy 356 KQLEEVSL-----ETRAQAMEKDQLAQLLEKDLLOKQVNEKIENL 396

Db 356 QVQVQLQVLLQLAHGGTLPGSITVPESENLSLMEKNQ--SLVEENEKLSRGLSEAAQT 413

Qy 397 TRML 400

Db 414 AQML 417

RESULT 9

US-10-334-143-8

; Sequence 8, Application US/10334143

; Publication No. US20040009549A1

; GENERAL INFORMATION:

; APPLICANT: GRIGORIEV, IGOR VYACHESLAVOVICH

; APPLICANT: SUDARSANAM, SUCHA

; TITLE OF INVENTION: METHOD FOR DETECTING REMOTE HOMOLOGUES AND NOVEL

; FILE REFERENCE: 038602/1543

; CURRENT APPLICATION NUMBER: US/10/334,143

; CURRENT FILING DATE: 2002-12-31

; PRIOR APPLICATION NUMBER: 60/343,169

; PRIOR FILING DATE: 2001-12-31

; NUMBER OF SEQ ID NOS: 207

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 8

; LENGTH: 1235

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-334-143-8

Query Match 33.8%; Score 694.5; DB 15; Length 1235;

Best Local Similarity 39.6%; Pred. No. 2.4e-55;

Matches 168; Conservative 66; Mismatches 145; Indels 45; Gaps 8;

Qy 6 VAVCVVRPLNSRESLG-ETAQVYWKTDNNVIYQVDSKSFNFRVPHGNETTKNVYEE 64

Db 13 VRVALRCPLVPKEISEGCMCLSFVPGEPQVVGTD--KSFTYDFVDPDTEQSEVENT 70

Qy 65 IAAPIIDSAIOGYNTIFAYGQTASGKTYTMWG-----SEDHLGVIPRAIHDFQKIK 117

Db 71 AVAPLKGIFKGYNATVLAYGQTGSGKTYSMGGAYTAQEENETVGVIPRVILQLLFXEID 130

Qy 118 KFPDREFLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNVVADLTVEVVTSEMA 177

Db 131 KKSDFEFTLVKSYLEIYNEETDILDCPSREKAQINREDPKKIKIVGLTEKTVLVALDT 190

Qy 178 LKWTKEKSHYGETKMNQSSRSHTIFRMLBSREKGFSPNCEGSKVSHNLNLDLAG 237

Db 191 VSCLEQGNNSRTVASTAMNSQSSRSHAFITISLEQRKSD--KNSFSRSLHLVDLAG 246

Qy 238 SERAAQTGAAGVRLKEGCNINRSFILGOVTKKLSGQVGGFINYRDSKLTTRILQNSLGG 297

Db 247 SERQKTKAEGDRLKEGININRGILLCLGNVISALGDDKGGFVFPYRDSKLTLLQLDSLGG 306

Qy 298 NPKTRIICTTPV--SFDETLTALQFASHTAKYMKNTPYVNEVSTDEALLKRYRKEIMDLK 355

Db 307 NSHTLMACVSPADSNLEETLNTLRVADRARKINKPIVN-----IDPQTAELNHLK 358

Qy 356 KQLEEVSL-----ETRAQAMEKDQLAQLLEKDLLOKQVNEKIENL 396

Db 356 QVQVQLQVLLQLAHGGTLPGSITVPESENLSLMEKNQ--SLVEENEKLSRGLSEAAQT 416

Qy 397 TRML 400

Db 417 AQML 420

RESULT 10

US-10-334-143-33

; Sequence 33, Application US/10334143

; Publication No. US20040009549A1

; GENERAL INFORMATION:

; APPLICANT: GRIGORIEV, IGOR VYACHESLAVOVICH

; APPLICANT: SUDARSANAM, SUCHA

; TITLE OF INVENTION: METHOD FOR DETECTING REMOTE HOMOLOGUES AND NOVEL

; FILE REFERENCE: 038602/1543

; CURRENT APPLICATION NUMBER: US/10/334,143

; CURRENT FILING DATE: 2002-12-31

; PRIOR APPLICATION NUMBER: 60/343,169

; PRIOR FILING DATE: 2001-12-31

; NUMBER OF SEQ ID NOS: 207

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 33

; LENGTH: 1237

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-334-143-33

Query Match 33.8%; Score 694.5; DB 15; Length 1237;

Best Local Similarity 38.8%; Pred. No. 2.4e-55;

Matches 167; Conservative 71; Mismatches 143; Indels 49; Gaps 10;

Qy 6 VAVCVVRPLNSRESLG-ETAQVYWKTDNNVIYQVDSKSFNFRVPHGNETTKNVYEE 64

Db 13 VRVALRCPLVPKEISEGCMCLSFVPGEPQVVGTD--KSFTYDFVDPDTEQSEVENT 70

Qy 65 IAAPIIDSAIOGYNTIFAYGQTASGKTYTMWG-----SEDHLGVIPRAIHDFQKIK 117

Db 71 AVAPLKGIFKGYNATVLAYGQTGSGKTYSMGGAYTAQEENETVGVIPRVILQLLFXEID 130

Qy 118 KFPDREFLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNVVADLTVEVVTSEMA 177

Db 131 KKSDFEFTLVKSYLEIYNEETDILDCPSREKAQINREDPKKIKIVGLTEKTVLVALDT 190

Qy 178 LKWTKEKSHYGETKMNQSSRSHTIFRMLBSREKGFSPNCEGSKVSHNLNLDLAG 237

Db 191 VSCLEQGNNSRTVASTAMNSQSSRSHAFITISLEQRKSD--KNSFSRSLHLVDLAG 246

Qy 238 SERAAQTGAAGVRLKEGCNINRSFILGOVTKKLSGQVGGFINYRDSKLTTRILQNSLGG 297

Db 247 SERQKTKAEGDRLKEGININRGILLCLGNVISALGDDKGGFVFPYRDSKLTLLQLDSLGG 306

Qy 298 NPKTRIICTTPV--SFDETLTALQFASHTAKYMKNTPYVNEVSTDEALLKRYRKEIMDLK 355

Db 307 NSHTLMACVSPADSNLEETLNTLRVADRARKINKPIVN-----IDPQTAELNHLK 358

Qy 356 KQLEEVSL-----ETRAQAMEKDQLAQLLEKDLLOKQVNEKIENL 396

Db 359 QVQVQLQVLLQLAHGGTLPGSINAESENLSLMEKNQ--SLVEENEKLSRGLSEAAQT 416

Qy 393 IENLRLMT 402

Db 417 AQMLERIL 426

RESULT 11

US-10-437-963-169150

; Sequence 169150, Application US/10437963

; Publication No. US20040123343A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.

; APPLICANT: Kovalic, David K.

; APPLICANT: Zhou, Yinhua

; APPLICANT: Cao, Yongwei

; APPLICANT: Wu, Wei

; APPLICANT: Boukharov, Andrey A.

; APPLICANT: Barbazuk, Brad

; APPLICANT: Li, Ping

Query Match 33.7%; Score 691.5; DB 16; Length 1029;
 Best Local Similarity 41.0%; Pred. No. 3.5e-55;
 Matches 171; Conservative 62; Mismatches 139; Indels 45; Gaps 11;
 ; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
 ; FILE REFERENCE: 38-21(53221)B
 ; CURRENT APPLICATION NUMBER: US/10/437,963
 ; CURRENT FILING DATE: 2003-05-14
 ; NUMBER OF SEQ ID NOS: 204966
 ; SEQ ID NO 169150
 ; LENGTH: 965
 ; TYPE: PRT
 ; ORGANISM: Oryza sativa
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: PAT_MRT4530_67597C.1.pap
 US-10-437-963-169150

Query Match 33.7%; Score 691.5; DB 16; Length 965;
 Best Local Similarity 40.8%; Pred. No. 3.2e-55; Indels 55; Gaps 10;
 Matches 166; Conservative 70; Mismatches 116;
 QY 1 AEEGAVAVCVVRPLNSR-----ESLGETAQVYKTDNNVIYQVDSKSFNDRVFH 53
 Db 17 AKERIMVSVRLPENGREAGDSCEISPTVMFRST---VPERAMPTAYTVDRVFG 73
 QY 54 GNETTKNVEIEAIIIDSAIQYNGTIFAYGOTAGSKTYTMGSEDHGLVPIRAIHDF 113
 Db 74 PDSSTQVVEEAGAKVALSVVSGINGSIFAYGOTSSGKTYTMTGITEY-----SVLDIY 127
 QY 114 QKIKKPPDFEFLRVSYMEIYNETITDLGCTQKMKPLIIRREDVNRNVVADLTEEVTY 173
 Db 128 DYTEKHPERFILFSAIETIYNEAVRDLL--SHDTPRLDDPEKTTVEKLTETFLRD 185
 QY 174 SEMALKWITKGEKSRHYGETKMNORSRSHITFRMILESRKGPSCGSKVSHNLV 233
 Db 186 KDLHRLNLLAVEAQRQIGETALNETSSRSQILR-----LNFV 223
 QY 234 DLASERAAQTGAAGVRLKEGCNINRSFLTGQVVKLSGQVGGFNYRDSKLTIRLON 293
 Db 224 DLASERASQTASAGVRLKEGSHINRSLLITLGVVRLSKGR--NGHIPYRDSKLTIRLOS 282
 QY 294 SLGNPKTRICTITPV--SFDETALQFASAKYMKNTPYVNEVSTDEALLKRYKEI 351
 Db 283 SLGNARTAICTMSPARSHIEQGRNTLLFATCAKEVNTNAQVNVVMSDKALVKHLQREL 342
 QY 352 MDLKKOLE---EVSLTRAQAM-BKD-----QLAKLEEKOLLQ 386
 Db 343 ERLQSEIKFPAPASCTTHAEALREKDAQIKKLEKQLKELMEERDTVK 389

RESULT 12
 US-10-311-642-2
 ; Sequence 2, Application US/10311642
 ; Publication No. US20040086878A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cytokinetics, Inc.
 ; APPLICANT: Beraud, Christophe
 ; APPLICANT: Freedman, Richard
 ; TITLE OF INVENTION: NOVEL MOTOR PROTEINS AND METHODS FOR THEIR
 ; FILE REFERENCE: 020552-001910US
 ; CURRENT APPLICATION NUMBER: US/10/311,642
 ; CURRENT FILING DATE: 2003-09-29
 ; PRIOR FILING DATE: 09/597,602
 ; NUMBER OF SEQ ID NOS: 4
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 2
 ; LENGTH: 1029
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: HsKif17 amino acid sequence
 ; DESCRIPTION OF ARTIFICIAL SEQUENCE: HsKif17
 US-10-311-642-2

Query Match 33.7%; Score 691.5; DB 16; Length 1029;
 Best Local Similarity 41.0%; Pred. No. 3.5e-55;
 Matches 171; Conservative 62; Mismatches 139; Indels 45; Gaps 11;
 QY 5 AVAVCVVRPLNSRREESLGETAQVYKTD-----NNVIYQVDSKSFNDRVPHGNE 56
 Db 5 AVKVVVRCRPMNQREEL--RCQPVVTVDCAQAQCCIQNPGGADEPPKQFTFDGAYVHDH 62
 QY 57 TTKNVEIEAIIIDSAIQYNGTIFAYGOTAGSKTYTMGSED---HLGVIPRAIHDF 113
 Db 63 VTEQIYNEIAYPLVEGVTEGYNGTIFAYGOTSGSKSFTMQGLPDDPPSQRGIIPRAEHVF 122
 QY 114 QKIKKPPDFEFLRVSYMEIYNETITDLGCTQKMKPLIIRREDVNRNVVADLTEEVTY 173
 Db 123 ESQQAENKFLVRASYLEIYNEVDRLDGADTKQK--LELKEHPEKGVVYKGLSMETVHS 181
 QY 174 SEMALKWITKGEKSRHYGETKMNORSRSHITFRMILESR---EKEPSCGSKVSHL 230
 Db 182 VAQCEHIMETGWNRSVGVTLNMKDSRSHISFTISIEMSAVDERG-----KDLHLAGKL 236
 QY 231 NLVDLAGSRAAQTGAAGVRLKEGCNINRSFLTGQVVKLSGQVGGFNYRDSKLTIRI 290
 Db 237 NLVDLAGSRSQSKTGATGERLKEATKINLSLALGNVISALVDGRC--KHVPYRDSKLTIRL 295
 QY 291 LQNSLGNPKTRICTITPV--SFDETALQFASAKYMKNTPYVNEVSTDEALLKRYR 348
 Db 296 LQDSLGNKTYLTMVCLSPADNNYDETLSTLYANRAKNIIRKNPRINEDPKD--ALLREYQ 354
 QY 349 KEIMDLKKOLEVSLTRAQAMEKDLQALL-----BEKOLLOKQVNEKIE 394
 Db 355 EEIKKLKAIL-----TQMSFSSLSALLSRQPPDPVQVBEKLLPQVPIQHDME 403

RESULT 13
 US-10-116-712-664
 ; Sequence 664, Application US/10116712
 ; Publication No. US20030194764A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Bangur, Chaitanya S.
 ; APPLICANT: Switzer, Ann
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
 ; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
 ; FILE REFERENCE: 210121.568
 ; CURRENT APPLICATION NUMBER: US/10/116,712
 ; CURRENT FILING DATE: 2002-04-07
 ; NUMBER OF SEQ ID NOS: 670
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 664
 ; LENGTH: 1232
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-116-712-664

Query Match 33.3%; Score 684.5; DB 14; Length 1232;
 Best Local Similarity 39.6%; Pred. No. 2.1e-54;
 Matches 168; Conservative 66; Mismatches 145; Indels 45; Gaps 9;
 QY 6 VAVCVVRPLNSRRESLG-ETAQVYKTDNNVIYQVDSKSFNDRVPHGNETTKNVEE 64
 Db 10 VRVALRCRPLVPKEISEGCQCLSEVPGEQPVVGTD--KSTYDFVDFDPSQEVEFNT 67
 QY 65 IAAPIIDSAIQYNGTIFAYGOTAGSKTYTMG-----SEDHIGVIPRAIHDFQKIK 117
 Db 68 AVAPLKGFKGNATVLAAYGQTSKTYSMGAYTAQEENBPTVGVIPRVITQLLFKEID 127
 QY 118 KFPDFEFLRVSYMEIYNETITDLGCTQKMKPLIIRREDVNRNVVADLTEEVTYSEMA 177
 Db 128 KKSDFEFTLVKSYLEIYNEEILDLCPREKAIINIREDPKGIKIVGTEKTVLVALDT 187
 QY 178 LKWTIKGEKSRHYGETKMNORSRSHITFRMILESRKGPSCGSKVSHNLVLDLAG 237
 Db 188 VSLQEGNNRSTVASTAMNSQSSRSHAFIT---SLEQKKSKDKNSSPR--SKLHLVDLAG 243

```

QY 238 SERAQTGAAGVRLKEGCNINRSFILGQVTKKLDGQGVGFNRYRDSKLTILQNSLGG 297
Db 244 SERQKTKAEGDRLKEGINRGLCLGNVISALGDDKGGFAPYRDSKLTLLQDSLGG 303
QY 298 NPKTRIICITIPV--SFDETLTALQFASHTAKYKNTPVYNEVSTDEALLKRYRKEIMDLK 355
Db 304 NSHTLMACVSPADSNLEETINTLYRADRKIKNKPVN-----IDPQTAEINHLK 355
QY 356 KQLEBESVL-----ETRAQAMEKDQLAQLLEEKDLLOKVONEKIENL 396
Db 356 QVQVQQLQVLLQAHGGTLPGSTITVPSENIQSLMEKNQ--SLVEENEKLSRGLSEAGQT 413
QY 397 TRML 400
Db 414 AQML 417

RESULT 14
US-10-116-712-669
; Sequence 669, Application US/10116712
; Publication No. US20030194764A1
; GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya S.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.568
; CURRENT APPLICATION NUMBER: US/10/116,712
; CURRENT FILING DATE: 2002-04-07
; NUMBER OF SEQ ID NOS: 670
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 669
; LENGTH: 1232
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-116-712-669

Query Match 33.3%; Score 684.5; DB 14; Length 1232;
Best Local Similarity 39.6%; Pred. No. 2.1e-54;
Matches 168; Conservative 66; Mismatches 145; Indels 45; Gaps 9;

QY 6 VAVCVVRVPLNSRESLG-ETAQVYWKTDNNVIYQVDSKSFNPRFVHGNETHKNNVYE 64
Db 10 VRVALRCEPLVPKEISECQMLCFVPGEPQVVGTD--KSFYDFVDPSTEQBEVENT 67
QY 65 IAAPIIDSAIQYNGTIPAYQTASGKTYTWMG-----SEDLGLVTPRAIHDFQKIK 117
Db 68 AVAPLILKGVFKGNATVLAQYQSGKTYSGGAYTAQENEPVGVIPRVLIQLLFKEID 127
QY 118 KFPDPREFLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNVVADLTVEVVYTGEMA 177
Db 128 KKSDFEFTLVSYLEIYNEEILDLCPREKAQINIREDPKEGKIVCLTEKTVLVALDT 187
QY 178 LKWTITKEKSRHYGETKMNQSSSHITFRLMLESREKGEPSNCEGSGVKVSHMLVDLAG 237
Db 188 VSCLEQGNKRTVASTANQSSSHAIPTI--SLEQGGKSKDNSSFR-SKLHLVDLAG 243
QY 238 SERAQTGAAGVRLKEGCNINRSFILGQVTKKLDGQGVGFNRYRDSKLTILQNSLGG 297
Db 244 SERQKTKAEGDRLKEGINRGLCLGNVISALGDDKGGFAPYRDSKLTLLQDSLGG 303
QY 298 NPKTRIICITIPV--SFDETLTALQFASHTAKYKNTPVYNEVSTDEALLKRYRKEIMDLK 355
Db 304 NSHTLMACVSPADSNLEETINTLYRADRKIKNKPVN-----IDPQTAEINHLK 355
QY 356 KQLEBESVL-----ETRAQAMEKDQLAQLLEEKDLLOKVONEKIENL 396
Db 356 QVQVQQLQVLLQAHGGTLPGSTITVPSENIQSLMEKNQ--SLVEENEKLSRGLSEAGQT 413
QY 397 TRML 400
Db 414 AQML 417

```

```

RESULT 15
US-10-408-765A-1664
; Sequence 1664, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1664
; LENGTH: 672
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-1664

Query Match 33.2%; Score 682.5; DB 16; Length 672;
Best Local Similarity 47.3%; Pred. No. 1.3e-54;
Matches 158; Conservative 55; Mismatches 106; Indels 15; Gaps 8;

QY 37 IYQVDSG-----KSFNFDVRFHGNETHKNNVYEBIAPIIDSAIQYNGTIPAYQTASGKT 92
Db 18 VHKTDSSNEPPKTFDTFVGPESKQLDVYNTLTPRIIDSVLEGYNGTIFAYGTGTGKT 77
QY 93 YTMGSE---DHLGVIPRAIHDFQKIKFP-DREFLLRVSYMEIYNETITDLCGTQK 148
Db 78 FTMEGVRAIPELRLGILPNSFAHIFGHIAKAEGRTRFLRVSYLEIYNEVEVDLL-GKQQT 136
QY 149 KPLIREDVNRNVVADLTVEVVYTSSEMALKWITKGEKSRHYGETKMNQSSSHITFIRM 208
Db 137 QRLVKERPDVGVYIKDLISAYVNVNADMDRIMTLGHKNSVSGATNMNHSRSHAIPTI 196
QY 209 ILESREKGEPSNCEGSGVKVSHNLVDLAGSERAAQTGAAGVRLKEGCNINRSFILGQVI 268
Db 197 TIECEKIDGNMH--VRMGKHLVDLAGSERQAKTGATQRLKEATKINLSLTGNVI 254
QY 269 KKLSDGQGVGFNRYRDSKLTILQNSLGGNPKTRIICITIPV--SFDETLTALQFASHTAK 326
Db 255 SALVDGK-STHVPYRNSKLTLLQDSLGGNSKTMNCANIGADYNYDETISTLYANRAK 313
QY 327 YMKNTPVYNEVSTDEALLKRYRKEIMDLKQLEE 360
Db 314 NIKNKARINEDPKD-ALLRQFQKEIEBELKKLEE 346

Search completed: July 29, 2004, 10:06:27
Job time : 22.1784 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 09:28:21 ; Search time 6.03629 Seconds
(without alignments)
6437.961 Million cell updates/sec

Title: US-10-045-631b-88_COPY_2_405
Perfect score: 2053
Sequence: 1 AEEGAVAVCVVRPLNSREE.....LQKVQNEKIENLRLMTSS 404

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR.78.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2053	100.0	2663	1 S28261	centromere protein
2	1468.5	71.5	2954	2 T14156	kinesin-related pr
3	872	42.5	823	2 T52425	kinesin-like prote
4	836.5	40.7	888	2 D96619	protein T30E16.9 [
5	754	36.7	1459	2 T30196	kinesin motor prot
6	723	35.2	701	1 B44259	kinesin-related pr
7	713	34.7	747	1 A57107	kinesin-related pr
8	712.5	34.7	742	1 S58691	kinesin-related pr
9	708.5	34.5	786	2 A53939	kinesin homolog KH
10	702	34.2	699	1 S38982	kinesin-related pr
11	694	33.8	1121	2 T06065	kinesin-related pr
12	693	33.8	1231	2 A54803	microtubule-associ
13	692	33.7	909	2 H86350	hypothetical prote
14	687	33.5	1226	2 T51617	hypothetical prote
15	686.5	33.4	581	2 F84599	kinesin-like prote
16	683	33.3	968	2 T45746	probable kinesin h
17	678	33.0	1130	2 T21134	hypothetical prote
18	677	33.0	932	2 T49235	hypothetical prote
19	667	32.5	1225	2 A56514	kinesin-like prote
20	662.5	32.3	1066	1 A48669	chromokinesin - ch
21	659.5	32.1	1263	2 T13465	kinesin-related pr
22	656	32.0	672	2 S54351	hypothetical prote
23	652	31.8	975	1 A31497	kinesin osm-3 - Ca
24	651	31.7	1254	2 T18277	kinesin heavy chai
25	649	31.6	1921	2 T13827	kinesin-73 - fruit
26	648.5	31.6	793	2 JCS831	kinesin-related pr
27	647.5	31.5	834	2 T06055	hypothetical prote
28	645	31.4	1031	1 A38713	kinesin heavy chai
29	644	31.4	1076	2 B84687	probable kinesin-l

30	637	31.0	967	1 A35075	kinesin heavy chai
31	634.5	30.9	1032	2 I38510	neuronal kinesin h
32	629.5	30.7	963	1 A41919	kinesin heavy chai
33	627	30.5	1027	2 S37711	kinesin heavy chai
34	625	30.4	784	1 A55236	kinesin-related pr
35	622.5	30.3	1388	2 T30135	KLP2 protein - Afr
36	622	30.3	843	2 S44868	kinesin heavy chai
37	622	30.3	1576	2 T29237	hypothetical prote
38	622	30.3	1584	1 UN0114	kinesin-related pr
39	622	30.3	1584	2 T15822	kinesin-like prote
40	621	30.2	294	2 S38983	kinesin-related pr
41	620	30.2	554	2 T50118	kinesin-related pr
42	620	30.2	857	2 E84600	probable kinesin h
43	616	30.0	968	2 T51933	kinesin motor prot
44	613	29.9	1022	2 E84792	probable kinesin h
45	610.5	29.7	881	2 I84737	kinesin heavy chai

ALIGNMENTS

RESULT 1

S28261
centromere protein E - human
N:Alternate names: centromere 312K protein; kinesin-related protein CENP-E
C:Species: Homo sapiens (man)
C:Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 19-Jan-2001
C:Accession: S28261
R:Yen, T.J.; Li, G.; Schaar, B.T.; Szilak, I.; Cleveland, D.W.
Nature 359, 536-539, 1992
A:Title: CENP-E is a putative kinetochore motor that accumulates just before mitosis.
A:Reference number: S28261; MUID:93024922; PMID:1406971
A:Accession: S28261
A:Molecule type: mRNA
A:Residues: 1-2663 <YEN>
A:Cross-references: EMBL:Z15005; NID:g29864; PIDN:CAA78727.1; PID:g29865
C:Genetics:
A:Gene: GDB:CENPE
A:Cross-references: GDB:361164; OMIM:117143
A:Map position: 4q24-4q25
C:Superfamily: centromere protein E; kinesin motor domain homology
C:Keywords: ATP; coiled coil; microtubule binding; mitosis; nucleotide binding; P-loop
F:7-335/Domain: kinesin motor domain homology <KMT>
F:86-93/Region: nucleotide-binding motif A [P-loop]
F:486-2183/Domain: coiled coil #status predicted <COI>
F:92/Binding site: ATP (lys) #status predicted

Query Match	100.00%;	Score 2053;	DB 1;	Length 2663;
Best Local Similarity	100.0%;	Pred. No. 4.5e-125;		
Matches 404;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	AEEGAVAVCVVRPLNSRESLGETAQVYWKTDNNVIYQVDGSKSFNDFRVFHNETTKN	60	
Db	2	AEEGAVAVCVVRPLNSRESLGETAQVYWKTDNNVIYQVDGSKSFNDFRVFHNETTKN	61	
QY	61	VVEEIAAPIIDSALQVNGTIFAYGQTASGKTYTMGSEDHGLGVIPRAIHDFQIKIKFP	120	
Db	62	VVEEIAAPIIDSALQVNGTIFAYGQTASGKTYTMGSEDHGLGVIPRAIHDFQIKIKFP	121	
QY	121	DREFFLLRVSYMEIYNETITDLCGTQKMKPLIIRBDVNRNVYVADLTETVVYTSSEALKW	180	
Db	122	DREFFLLRVSYMEIYNETITDLCGTQKMKPLIIRBDVNRNVYVADLTETVVYTSSEALKW	181	
QY	181	ITKGEKSRHYGETKMQNQRSSRSHITFRMILESRKEGSPNCEGSKVSHNLVDLAGSR	240	
Db	182	ITKGEKSRHYGETKMQNQRSSRSHITFRMILESRKEGSPNCEGSKVSHNLVDLAGSR	241	
QY	241	AAQTGAAGVRLKEGCNINRSILFGQVIKLSDGVGGFINVRDSKLTILQNSLGNPK	300	
Db	242	AAQTGAAGVRLKEGCNINRSILFGQVIKLSDGVGGFINVRDSKLTILQNSLGNPK	301	
QY	301	TRIICTITPVSFDEFTLALQFASTAKYMKNTPVYNEVSTDEALLKRYRKEIMDLKQLEE	360	

Db 302 TRIICTTPVSFDETLTALQFASAKYMKNTPYVNEVSTDEALLKRYRKEIMDLKKOLEE 361

Qy 361 VSLFTRQAMKQDLAQLEKDLLOKQVNEKIENLTMLVTSS 404

Db 362 VSLFTRQAMKQDLAQLEKDLLOKQVNEKIENLTMLVTSS 405

RESULT 2

T14156

kinesin-related protein - African clawed frog

C:Species: Xenopus laevis (African clawed frog)

C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-May-2000

C:Accession: T14156

R:Wood, K.W.; Skowicz, R.; Goldstein, L.S.; Cleveland, D.W.

Cell 91, 357-366, 1997

A:Title: CENP-E is a plus end-directed kinetochore motor required for metaphase chromosome

A:Reference number: Z17893; MUID:98028574; PMID:9363944

A:Accession: T14156

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-2954 <WOO>

A:Cross-references: EMBL:AF027728; NID:g2586070; PID:g2586071; PIDN:AAC60300.1

C:Genetics:

A:Gene: XENP-E

C:Superfamily: centromere protein E; kinesin motor domain homology

Query Match 71.5%; Score 1468.5; DB 2; Length 2954;

Best Local Similarity 70.9%; Pred. No. 5.4e-87;

Matches 290; Conservative 52; Mismatches 60; Indels 7; Gaps 3;

Qy 1 AERGAVAVCVVRPLNSRESLGETAQVYWKTDNNVIYQVDSKSNFDRVFNHNETKN 60

Db 2 SEGDAVKVCVRPLIQREQ--GDQANLQWKAGNNTISQVDTGKSFNDFRVFNSHSTSQ 59

Qy 61 VYEEIAAPIIDSAIQVNGTIFAYGQTASGKTYTMGSEDLHGVIPRAIHDIQKIKFP 120

Db 60 IYGEIAPVIRLSALQVNGTIFAYGQTSSGKTYTMGTSPNSLGIIPQAIQVFKIIEIP 119

Qy 121 DRPEFLVSYMEIYNETITDLCQTKWKPLIREDVRNVYVADLTVEEVVYSEMAKW 180

Db 120 NREFFLVSYMEIYNETVTKLLCDRRKKPLREIDRNRNVYVADLTVEELVMPVPIQW 179

Qy 181 ITGKEGRHGYETQKQSSRSHTIFRMILESREKGEPS--NCEGSVKYVSHLMLVDLAG 237

Db 180 IKGKRNHGYETQKNDHSSHTIFRMIVESRDRNDPTNSGCDGAVVYSHLMLVDLAG 239

Qy 238 SERAAQTCAAGVRLKEGCNINRSFILGQVKKLSGQVGFYNYRDSKLTRELQNSLGG 297

Db 240 SERASQTCAEGVRLKEGCNINRSFILGQVKKLSGQAGFYNYRDSKLTRELQNSLGG 299

Qy 298 NPKTRIITTPVSFDETLTALQFASAKYMKNTPYVNEVSTDEALLKRYRKEIMDLKKQ 357

Db 300 NAKTVIITTPVSFDETLTALQFASAKYMKNTPYVNEVSTDEALLKRYRKEIMDLKKQ 359

Qy 358 LE--EVSLFTRQAMKQDLAQLEKDLLOKQVNEKIENLTMLVTSS 404

Db 360 LENLESSETQAMAKKEHTQLAEIKQLKHEREDRIWHLNTIVVASS 408

RESULT 3

T52425

kinesin-like protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 24-Oct-2000 #sequence_revision 24-Oct-2000 #text_change 24-Oct-2000

C:Accession: T52425

R:Kato, A.; Suzuki, M.; Kuwahara, A.; Ooe, H.; Higano-Inaba, K.; Komeda, Y.

Gene 239, 309-316, 1999

A:Title: Isolation and analysis of cDNA within a 300 kb Arabidopsis thaliana genomic region

A:Reference number: Z25171

A:Accession: T52425

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-823 <KAT>

A:Cross-references: EMBL:AB028468; PIDN:BAA88112.1

C:Genetics:

A:Gene: ZCF125

Query Match 42.5%; Score 872; DB 2; Length 823;

Best Local Similarity 47.6%; Pred. No. 6.1e-49;

Matches 204; Conservative 58; Mismatches 123; Indels 44; Gaps 9;

Qy 6 VAVCVVRPLNSRESLGETAQVYWKTDNNVIYQVDSKSNF--FDRVPHGNE 56

Db 4 ICVAVVRVP-----PAPENGASLWKVEDN--RISLHKSLDPTPTTASHAFDHFDESS 54

Qy 57 TTKNVYEEIAPIIDSAIQVNGTIFAYGQTASGKTYTMGSEDLHGVIPRAIHDIQKI 116

Db 55 TNASVIELTKDIIHAAVEGFGTAFAYGQTSSGKTFMTGSETDPGLIIRSRVDRVFERI 114

Qy 117 KKFDPREFLIRVSYMEIYNETITDLCQTKWKPLIREDVRNVYVADLTVEEVYVTS 176

Db 115 HMSDREFLIRVSYMEIYNEEINDLL--AVENQLQIHEHLRGVFAGLKEEIVSDAQ 172

Qy 177 ALKWTGKSHRYGETQKQSSRSHTIFRMILESREKGEPSNCEGSKVSHLMLVDLA 236

Db 173 ILKLDSGEVNRHFGETNMNVHSSRSHTIFRMVIESR--GKNSSSDAIRVSVNLVDLA 230

Qy 237 GSERAAQTCAAGVRLKEGCNINRSFILGQVKKLSG--QVGGFYNRDSKLTRELQNSL 295

Db 231 GSERIAKTGAGVRLQEGKYINKSLMILGNVINKLSGSTKLRAHIPYRDSKLTRELQPAL 290

Qy 296 GGNPKTRIITTPVVS--FDETLTALQFASAKYMKNTPYVNEVSTDEALLKRYRKEIMD 353

Db 291 GGNAKTCTIITTAPEEHIEESKGTLOFASRAKRNITCAQVNEILTDALLKKQLEIEE 350

Qy 354 LKQLEEVSLFTRQ-----AMEKQDLAQLEKDLLOKQVNE-----KIE 394

Db 351 LEMKLGSHAEVLEQILNLSNMLKVECEKTKQLEBEKQKQEQENCIEKQMQMKIE 410

Qy 395 NLTRMLVTS 403

Db 411 NLNFTVNS 419

RESULT 4

D96619

protein T30E16.9 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: D96619

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: D96619

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-888 <STO>

A:Cross-references: GB:AE005173; NID:g8778739; PIDN:AAF79747.1; GSPDB:GN00141

C:Genetics:

A:Gene: T30E16.9

A:Map position: 1

Query Match 40.7%; Score 836.5; DB 2; Length 888;

Best Local Similarity 43.5%; Pred. No. 1.4e-46;

Matches 204; Conservative 58; Mismatches 126; Indels 81; Gaps 10;

Qy 6 VAVCVVRPLNSRESLGETAQVYWKTDNNVI-----Y 38


```

Db      4 ICVAVKVRP-----PAPENGASLWKVEDNRNLSLHKSLDTPITTAASHAFVSGISISTDLI 57
QY      39 QVDSKSFNF-----DRVFGHNETTKNVYEEIAPIIDSATQGYNGTIFAYGOTA 88
Db      58 EIVSLFLFSGVVYFFLLADHVDFESSNASVYELTWDIIHAAVEGNGTAFAYGOTS 117
QY      89 SKTYTMMGSDHLGVIPRAIHDI FQIKKFPDREFLLRVSMEIYNEITIDLLCGTQRM 148
Db      118 SKTFTMTGSETDPIIRSRVDPFERIHMISDREFLIRVSMEIYNEINDLL--AVEN 175
QY      149 KPLIREDNRNVVYADLTVEVYVISEMALKWITKGEKSRHGETVMQNRSSRSHITFRM 208
Db      176 QRLQIHEHLRGVVFAGLKEEVISDAEQILKIDSGEVNRFHGETMNMVHSSRSHITFRM 235
QY      209 -----ILSREKGEPSNCGSVKVSGLNLDLAGSRAAQTGAAGVRLKEGCN 256
Db      236 VFRSYERDLLLVIESR--GKNSSSDAIRVSVNLNLDLAGSERIAKTGAGVRVQEGKY 293
QY      257 INRSFLIQGVIKKLSDG-QVGGFNYRDSKLTILQNSLGGNPKTRIICTITPVS--FD 313
Db      294 INKSLMILGNVINKLSDSSTKLRAHIPYRDSKLTILQPALGNAKTCIITIAPEEHIE 353
QY      314 EFLTALQFASAKYMKNTFYVNEVSTDEALLKRYKREIMDLKQLEVSLETRAQ----- 368
Db      354 ESKGTQLQFASRAKRTNCQAQVNEIILTDAALKRKQLEIEELRMKLGQSHAEVLEQIILML 413
QY      369 -----AMEKDQLAQLLEEKDLOKQVNE-----KIENLTRMLVTS 403
Db      414 SNQMLKYELECERLTKQLEEEKQKEQENYCKEQQMKIENLNNFTNS 462

RESULT 5
T30196
Kinesin motor protein 1 - smut fungus (Ustilago maydis)
C:Species: Ustilago maydis (corn smut)
C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 11-May-2000
C:Accession: T30196
R:Lehmler, C.; Steinberg, G.; Snetselaar, K.M.; Schliwa, M.; Kahmann, R.; Bolker, M.
EMBO J. 16, 3464-3473, 1997
A:Title: Identification of a motor protein required for filamentous growth in Ustilago
A:Reference number: Z20770; MUID:97361828; PMID:9218789
A:Accession: T30196
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1459 <LEH>
A:Cross-references: EMBL:U92844; NID:g2062749; PID:g2062750; PIDN:AAB63336.1
C:Genetics:
A:Gene: kin1
C:Function:
A:Description: required for filamentous growth in Ustilago maydis

Query Match      36.7%; Score 754; DB 2; Length 1459;
Best Local Similarity 38.2%; Pred. No. 6.2e-41;
Matches 190; Conservative 70; Mismatches 127; Indels 110; Gaps 14;

QY      6 VAVCVVRPLNSREBSLGETAQVYKTDN--NVIIYQVD----- 41
Db      237 VVVCVRMRP--SRASSSDSEASV-WNCDSKRNRIPTTEHHPALAKRTTSSERAGAGASIA 293
QY      42 -----GSKSFNPDVFNHNETTKNVYEEIAPIIDSATQGYNGTIFAYGQ 86
Db      294 AAPSSHDLDHEDPTSTSYHFQDKLITGAQTDDMYHSHIAPVVRAAVEGYNGTVFAYGQ 353
QY      87 TASGKTYTMMGSEDHLGVIPRAIHDI FQIKKFPDPREFLLRVSMEIYNEITIDLLCGTQ 146
Db      354 TGSCKTHTMGSDDAPGVIIPRAVEGIQFMIXDEPDREFLLRVSYLEIYNETLKKLLA--- 410
QY      147 KKKPL-----IREDNRNVVYADLTVEEV 171
Db      411 PLPPIGTSGSLQTTDRPASPIKGGSSHAQSQSCITRIIEIDQKSRVITGLUREIV 470
QY      172 YTEMALKWIYKGEKSRHYGETKMNQNRSSRSHITFRMILSREKGEPSNCGSVKVSGLN 231

```

```

Db      471 TDANTVLCIORQDQDERHVGAIDWNRSSRSCHVCVOLTIESRSPASASKE--VRISQLN 528
QY      232 IYDLAGSRAAQTGAAGVRLKEGCNINRSFLILQGVIKKLSGQVGG--FINYRDSKLTR 289
Db      529 LIDLAGSRAA---SQAERKEGAFINKLSLLTIGTVIGKLTPEVNGDAHPIYRDSKLTR 585
QY      290 ILQNSLGGNPKTRIICTITPVS--PDETLTALQFASAKYMKNTFYVNEVSTDEALLKRY 347
Db      586 ILQTSLSGNARLAVICTISPDTEHANEILTSLKFGKRCKLVTATTAKGTAMDKALLQKY 645
QY      348 RKEIMDLKKQLEB-----VSLERAQAMEK-DQLAQLLE-----EKDLLQKQV 389
Db      646 RKELDALARAKLKEANGSPNPNEVMTIVSAESKESQKLDQLNQKEAAQREVEDMQKR 705
QY      390 NE---KIENLTRMLVTS 403
Db      706 SHLKAQIEHLTRILTS 722

RESULT 6
B44259
Kinesin-related protein KIF3A - mouse
C:Species: Mus musculus (house mouse)
C>Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 19-Jan-2001
C:Accession: B44259; S27872
R:Aizawa, H.; Sekine, Y.; Takemura, R.; Zhang, Z.; Nangaku, M.; Hirokawa, N.
J. Cell Biol. 119, 1287-1296, 1992
A:Title: Kinesin family in murine central nervous system.
A:Reference number: A44259; MUID:93077686; PMID:1447303
A:Accession: B44259
A:Molecule type: mRNA
A:Residues: 1-701 <AIZ>
A:Cross-references: EMBL:D12645; NID:g220469; PID:BAA02166.1; PID:g220470
A:Experimental source: brain
A>Note: sequence extracted from NCBI backbone (NCBIP:118911)
C:Complex: heterodimer with KIF3B (PIR:A57107); the KIF3A/3B heterodimer associates with
C:Function:
A:Description: KIF3 complex is a motor protein that provides anterograde fast axonal tra
A:Superfamily: kinesin-related protein KIF3; kinesin motor domain homology
C:Keywords: ATP; coiled coil; heterodimer; heterotrimer; microtubule binding; nucleotide
F:1-368/Domain: head globular #status predicted <HGI>
F:15-351/Domain: kinesin motor domain homology <KMOT>
F:100-107/Region: nucleotide-binding motif A (p-loop)
F:369-595/Domain: helical rod #status predicted <ROD>
F:600-701/Domain: tail globular #status predicted <TGL>
F:106/Binding site: ATP (lys) #status predicted

Query Match      35.2%; Score 723; DB 1; Length 701;
Best Local Similarity 44.4%; Pred. No. 2.4e-39;
Matches 178; Conservative 66; Mismatches 137; Indels 20; Gaps 10;

QY      6 VAVCVVRPLNSREBSL--GETAQVYKTDNVIYQVDGS----KSFNPDVFNHNETTK 59
Db      15 VKVVVRCPLNREKREKSMCYQAVSDENRGITVHKTDSSNEPEPKTFTFTDVFVGESQL 74
QY      60 NVYEIEAAPIIDSATQGYNGTIFAYGOTASGTITVTMGSE---DHLGVIPRAIHDI FQKI 116
Db      75 DVYNLTARPIIDSLEGGNGTIFAYGQGTGKTFTMEGVRAVPGILRGVIPNSFAHIFGHI 134
QY      117 KKFP--DREFLLRVSMEIYNETITDLCGTQKMKPLIREDNRNVVYADLTVEEVYTS 175
Db      135 AKAESGDTFLVRVSYLEIYNEEVRDLL-GKQOTQRLVKERPDGVYIKDLSAYVNNAD 193
QY      176 MALKWITKGEKSRHYGETKMNQNRSSRSHITFRMILSREKGEPSNCGSVKVSGLNLDL 235
Db      194 DMDRIMTLGHKNRSGVATNNNEHSSRSHAITTITIECKEKGVDGNMH--VRMGKHLVDL 251
QY      236 AGSRAAQTGAAGVRLKEGCNINRSFLILQGVIKKLSGQVGGFINYRDSKLTRILQNSL 295
Db      252 AGSRQAKTGTATGQRLKEATKINLSLTGLNVISALVDGK-STHVPYNSKLTRILQDLS 310
QY      296 GGNPKTRIICTITPV--SFDETLTALQFASAKYMKNTFYVNEVSTDEALLKRYRKEIMD 353

```

Db 311 GGNKTMCMCANIGPADNYDNTISTLRYANRAKNKINAKARINEDPKD-ALLRQFQKEIEE 369
354 LKKQL---BEVSLTETRAQAMEKDQLAQLEEKLLQKVQNE 391
370 LKKKLREGESVSGDISGSEDEDEGELGEGEKRRRDQ 410

RESULT 7

A57107
Kinesin-related protein KIF3B - mouse
C:Species: Mus musculus (house mouse)
C>Date: 03-Nov-1995 #sequence_revision 03-Nov-1995 #text_change 19-Jan-2001
C:Accession: A57107
R;Yamazaki, H.; Nakata, T.; Okada, Y.; Hirokawa, N.
J. Cell Biol. 130, 1387-1399, 1995
A;Title: KIF3A/B: a heterodimeric kinesin superfamily protein that works as a microtubul
A;Reference number: A57107; MUID:96032268; PMID:7559760
A;Accession: A57107
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-747 <YAM>
A;Cross-references: GB:D26077; NID:g1060922; PIDN:BAA05070.1; PID:g1060923
A;Experimental source: Brain
C;Complex: heterodimer with KIF3A (PIR:B44259); the KIF3A/3B heterodimer associates with
C;Function:
A;Description: KIF3 complex is a motor protein that provides anterograde fast axonal tra
C;Superfamily: kinesin-related protein KIF3; kinesin motor domain homology
C;Keywords: ATP; coiled coil; heterodimer; heterotrimer; microtubule binding; nucleotide
F;1-363/Domain: head globular #status predicted <HGL>
F;10-346/Domain: kinesin motor domain homology <KMT>
F;96-103/Region: nucleotide-binding motif A (P-loop)
F;364-592/Domain: helical rod #status predicted <ROD>
F;594-747/Domain: tail globular #status predicted <TGL>
F;102/Binding site: ATP (Lys) #status predicted

Query Match 34.7%; Score 713; DB 1; Length 747;
Best Local Similarity 41.4%; Pred. No. 1.2e-38;
Matches 177; Conservative 69; Mismatches 144; Indels 38; Gaps 10;
Qy 5 AVAVCVVRPLNSREESLG-----ETAAQVYKTDNNVIYQVDSKSPNDFRVPHGN 55
Db 9 SVRVVRCRPMNGEKAASYDKVDVVLQGVSVKPKGTSHEM--PKTTFDAVYDWN 66
Qy 56 ETTKNVVEEIAPIIDSAGIYNGTIFAYGQTASGKTYTM---MGSEDLGVIPRAIHI 112
Db 67 AKQELYDETRPLVDVSLQFNGTIFAYGQTGKTVMGVRGDPKRGVIPSFDHI 126
Qy 113 FQTKKPPDRFLRVSMYIYNETITDLCGQKMKPLIREDVNRVYVADLTVEVY 172
Db 127 FTHSRNQQLVRASYLEIYQBEIRDLKQD--TKRLKRPDGTGVYVKDLSFVTK 185
Qy 173 TSEALKWITGKSRHYGTEKMNORSRSHITFRMILESEKGPSPNCEGSKVSHLNL 232
Db 186 SVKEIEHVNVGNQNSVGATNMHSSRSHAFVITTECEVG--LDGENHVRGKLN 243
Qy 233 VDLAGSERAQTGAAGVRLKEGCNINRSLFLGQVKKLSDQGVGFINRDSKLTRLQ 292
Db 244 VDLAGSEQAQTGAQGERLKEATKINLSLGNVISALVDGK-STHIPYRDSKLTRLQ 302
Qy 293 NSLGNPKTRIICTITPVSF--DETTLAQFASAKYKNTPYNEVSTDEALKRYKE 350
Db 303 DSLGNAKTVMVANMGPASYNVEETLTLLRYANRAKNKPNRVNEDPKD-ALLREFQEE 361
Qy 351 IMDLKKQLREVSLETRAQ-----AMEKQDLAQLEE-----KDLIQKVQNEKI 393
Db 362 IARLKAQLEKRSIGRRKREKREKREGGSGGGGEEEGEEGEGEDDKDDKYWRQOEKL 421
Qy 394 ENLIRMLV 401
Db 422 EIEKRAIV 429

RESULT 8

S58691
Kinesin-related protein KRP95 - sea urchin (Strongylocentrotus droebachiensis)
N;Alternate names: kinesin-2 chain B; KRP (85/95) 95K chain
C;Species: Strongylocentrotus droebachiensis
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-Jan-2001
C:Accession: S58691
R;Rashid, D.J.; Wedaman, K.P.; Scholey, J.M.
J. Mol. Biol. 252, 157-162, 1995
A;Title: Heterodimerization of the two motor subunits of the heterotrimeric kinesin, KR
A;Reference number: S58691; MUID:95404610; PMID:7674298
A;Accession: S58691
A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
A;Molecule type: mRNA
A;Residues: 1-742 <RAS>
C;Complex: heterotrimer of a 115K chain and two kinesin-related chains of 85K (PIR:S5899
C;Superfamily: kinesin-related protein KIF3; kinesin motor domain homology
C;Keywords: ATP; heterotrimer; microtubule binding; nucleotide binding; P-loop
F;9-345/Domain: kinesin motor domain homology <KMT>
F;95-102/Region: nucleotide-binding motif A (P-loop)
F;101/Binding site: ATP (Lys) #status predicted

Query Match 34.7%; Score 712.5; DB 1; Length 742;
Best Local Similarity 41.3%; Pred. No. 1.3e-38;
Matches 175; Conservative 65; Mismatches 137; Indels 47; Gaps 10;
Qy 6 VAVCVVRPLNSREESLGETAQVYKTDNNVIYQV-----DGSKSFNDFRVPHGNETT 58
Db 9 VKVVVRCRPMNSKEISQGHKRIIVEMDNKRGLEVNTNPKGPPGEPNKSFTDVTVDWNSKQ 68
Qy 59 KNYVEEIAAPIIDSAGIYNGTIFAYGQTASGKTYTMWG---SEDHLGVIPRAIHDIFQK 115
Db 69 IDLYDETRFLRVSLQFNGTIFAYGQTGKTFTMEGVRSNPELRGVIPNSPEHIFTH 128
Qy 116 IKKPPDRFLRVSMYIYNETITDLCGQKMKPLIREDVNRVYVADLTVEVYVISE 175
Db 129 IARTNQQLVRASYLEIYQBEIRDLAKDQK-KRLDLKERPDGTGVYVKDLSFVTSVK 187
Qy 176 MALKWITGKSRHYGTEKMNORSRSHITFRMILESEKGPSPNCEGSKVSHLNLVDL 235
Db 188 EIEHVTMTVGNNSRVSSTNMHSSRSHAFIITIESELGVDG--ENHIVRGKLNVDL 245
Qy 236 AGSERAQTGAAGVRLKEGCNINRSLFLGQVKKLSDQGVGFINRDSKLTRLQNSL 295
Db 246 AGSERQAQTGAQGERLKEATKINLSLGNVISALVDGK-SSHIPYRDSKLTRLQDSL 304
Qy 296 GGNPKTRIICTITPVS--FDTTLAQFASAKYKNTPYNEVSTDEALKRYKKEIMD 353
Db 305 GGNAKTVMVANMGPASYNFDETTTLRYANRAKNKPNKINEDPKD-ALLREFQEEISR 363
Qy 354 LKKQL-----EEVSLTETRAQAMEKDQLAQLEE--KDLLQKV 388
Db 364 LKQALDKGPSDRKKKRPQGGDDDDIEDTEEGEDME-----EEMYKESQQL 418
Qy 389 QNEK 392
Db 419 EEEK 422

RESULT 9

A53939
Kinesin homolog Khp1 - Chlamydomonas reinhardtii
C:Species: Chlamydomonas reinhardtii
C>Date: 07-Oct-1994 #sequence_revision 07-Oct-1994 #text_change 02-Feb-2001
C:Accession: A53939
R;Walther, Z.; Vaeshishta, M.; Hall, J.L.
J. Cell Biol. 126, 175-188, 1994
A;Title: The Chlamydomonas FLA10 gene encodes a novel kinesin-homologous protein.
A;Reference number: A53939; MUID:94299638; PMID:8027176
A;Accession: A53939
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-786 <WAL>
A;Cross-references: EMBL:L33697; NID:g497696; PIDN:AAA21738.1; PID:g497697

A;Note: authors translated the codon AAC for residue 753 as Asp

C;Genetics:

A;Gene: FLA10

C;Superfamily: unassigned kinesin-related proteins; kinesin motor domain homology

C;Keywords: ATP; coiled coil; nucleotide binding; P-loop

F;11-359/Domain: kinesin motor domain homology <KMOT>

F;97-104/Region: nucleotide-binding motif A (P-loop)

Query Match 34.5%; Score 708.5; DB 2; Length 786;

Best Local Similarity 44.4%; Pred. No. 2.5e-38;

Matches 167; Conservative 66; Mismatches 118; Indels 25; Gaps 9;

QY 5 AVACVVRPLNSREESLGTAQVYVKTNNVY----QVDGS---KSENFDRVFGNET 57

DB 10 SVKVVRCRPLNGKEKADGRSIRVDMVDAGQVKVRNPKADASEPPKAFPTDQVYDNQC 69

QY 58 TKNVVEEIAAPIIDSAIQGYNGTIFAYGQTASGKTYTMMGSEDH---LGVIPRAIHDFQ 114

DB 70 QRDVFDITAPRLIDSCIEGYNGTIFAYGQTGKSHTWEGKDEPPELRLGLIPNTRYVFE 129

QY 115 KI-KKFPDRFLRVSVMEIYNETITDLCGTQKMKPLIREDVNRNVYVADLTEEYVT 173

DB 130 IARDSGTKFELVRSSYLEIYNEVRDIL--GKDSHKMKELKESPDGYYVVKDLSQVCKN 188

QY 174 SEMALKWITKGEKSRHYGETKMQRSRSHITFRMILESREK-----GEPSNCEGS 224

DB 189 YEEMKVKLLAGKNRQVGATLMNQDSKSRSHSIFITIEIEKLSAAAKPFAKKDDSNH 248

QY 225 VKVSHLNLVDLAGSERAQAOTGAAGVRKKEGNCINRSFILGQVVKKLSGQGVGFINRYD 284

DB 249 VRVGLNLVDLAGSERQDKTGATGDLKEGKIKNLSTALGNVISALVDGK-SGHIPYRD 307

QY 285 SKLFRILQNSLGNPKTRIITCTTPV--SPDETALQFATAKYMKNTPYVNEVSTDEA 342

DB 308 SKLRLQLDSLGNTKTVWVANIGPADWNVDYETSTLYANRAKNIQNKPKINEDPKD-A 366

QY 343 LKRYRKIMDLKKQL 358

DB 367 MLRQFQEEIKLKBQL 382

RESULT 10

S38982

Kinesin-related protein KRP85 - sea urchin (Strongylocentrotus purpuratus)

N;Alternate names: kinesin-2 chain A; KRP (85/95) 85k Chain

C;Species: Strongylocentrotus purpuratus (purple urchin)

C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-Jan-2001

C;Accession: S38982

E;Cole, D.G.; Chinn, S.W.; Wedaman, K.P.; Hall, K.; Vuong, T.; Scholley, J.M.

Nature 366, 268-270, 1993

A;Title: Novel heterotrimeric kinesin-related protein purified from sea urchin eggs.

A;Reference number: S38982; PMID:8232586

A;Accession: S38982

A;Molecule type: mRNA

A;Residues: 1-699 <COL1>

A;Cross-references: EMBL:U16993; NID:g295245; PIDN:AAA16098.1; PID:g295246

A;Accession: S72551

A;Molecule type: protein

A;Residues: 2-5,'X',7-11,59-64,125-132,222-226,'X',228-230 <COL2>

C;Complex: heterotrimer of a 115k chain and two kinesin-related chains of 95k (PIR:S5869)

C;Superfamily: kinesin-related protein KIF3; kinesin motor domain homology

C;Keywords: ATP; heterotrimer; microtubule binding; nucleotide binding; P-loop

F;11-348/Domain: kinesin motor domain homology <KMOT>

F;97-104/Region: nucleotide-binding motif A (P-loop)

F;103/Binding site: ATP (Lys) #status predicted

Query Match

Best Local Similarity 34.2%; Score 702; DB 1; Length 699;

Matches 190; Conservative 65; Mismatches 131; Indels 92; Gaps 16;

QY 6 VAVCVVRPLNSREESLG-----ETAQVYVKTNNVYQVDGSKSFNDRVFHG 54

DB 11 VRVVRPLNSKETCGGKSVVKMDMGTVQV---TNPNA-PSGEPPKSTFDIVFAP 66

QY 55 NETTKNVYEIAAPIIDSAIQGYNGTIFAYGQTASGKTYTMMG--SDHL-GVIPRAIHD 111

DB 67 GAKQTDVYNGTARPIVDALIEGYNGTIFAYGQTGKTFTMEGVRSQPELRLGILPNSFAH 126

QY 112 IFQIKKFPDR-BFLRVSVMEIYNETITDLCGTQKMKPLIREDVNRNVYVADLTEEY 170

DB 127 IFGHIAKEQENRFLRVSVYLEIYNEVRDIL--GKQOQHRLEVKRERPDVGVYVVKDLSAFV 185

QY 171 VYVSEMALKWITKGEKSRHYGETKMQRSRSHITFRMILESREKGEPSNCEGSVKVSHL 230

DB 186 VNNADDMDRITMTLGNKRSVSGATNMNESSRSHAFTITLERSDMG--LDKEQHVRYGKL 243

QY 231 NLVDLAGSERAQAOTGAAGVRKKEGNCINRSFILGQVVKKLSGQGVGFINRYDSKLTRI 290

DB 244 HMVDLAGSERQDKTGATGDLKEGKIKNLSTALGNVISALVDGK-STHPIYRNSKLTRL 302

QY 291 LQNSLGNPKTRIITCTTPV--SPDETALQFATAKYMKNTPYVNEVSTDEALLKRYR 348

DB 303 LQDSLGNNAKTVMCANIGPAEYNYDETISTLYANRAKNIKNAKINEDPKD-ALLREFQ 361

QY 349 KEIMDLKKOLEE-----VSLETRAQAM 370

DB 362 KEIEELKKQISESGEGGLDDDESGSESGEAGEGGVKKRKNPKRKLSPKLSPEIMA-AM 420

QY 371 EK--DOLAQLLE-KDLL-----QKVQNEKIENLTRLV 401

DB 421 QKIDEEKKALEEKMDVVEDRNTVHRELQRESELHKAQDDQKILNEKLNIAQKKLI 478

RESULT 11

T06065

hypothetical protein F19H22.150 - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 31-Mar-2000

C;Accession: T06065

R;Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X

submitted to the Protein Sequence Database, March 1999

A;Reference number: 215184

A;Accession: T06065

A;Molecule type: DNA

A;Residues: 1-1121 <BEV>

A;Cross-references: EMBL:AL035679; GSPDB:GN00062; ATSP:F19H22.150

A;Experimental source: cultivar Columbia; BAC clone F19H22

C;Genetics:

A;Gene: ATSP:F19H22.150

A;Map position: 4

A;Introns: 139/2; 170/1; 200/1; 256/3; 284/3; 302/3; 322/3; 349/1; 371/3; 390/3; 425/3;

C;Superfamily: kinesin heavy chain; kinesin motor domain homology

F;99-494/Domain: kinesin motor domain homology <KMOT>

Query Match

Best Local Similarity 33.8%; Score 694; DB 2; Length 1121;

Matches 178; Conservative 87; Mismatches 131; Indels 96; Gaps 14;

QY 1 ABEGAVAVCVVRPLN-----SRRES-----IGETA---QVYVTKDN 34

DB 94 SERDSISVTVRPRPLRYARSDLAMDQIYCRDYSFHVDAIGVNSLLGQESFGIIPAIKDV 153

QY 35 NVI--YQVDGSKSFNDRVFNHNETKNVYEIAAPIIDSAIQGYNGTIFAYGQTASGKT 92

DB 154 DTLVRHEYNPLTAVAFDKVFGPQATTIDVDVAARPVVKAAMEGVNGVYFAVGVTSSTGKT 213

QY 93 YTM-----MGSEDLHGVIPRAIHDF 113

DB 214 HTMHVRVLKKNPYTPPSLFLQNTCANWDLFFNLVPLNYLLGQESFGIIPAIKDV 273

QY 114 QKIKKF-----PDREFLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNVYVADLT 167

DB 274 SIIQDVSGLNGTPGREFLLRVSYLEIYNEVINDLLDPTG--QNLVRRED-SQGTIVEGIF 330

QY 168 BEVVYVTSMALKWTIKGKSRHYGETKMQRSRSHITFRMILESREKGEPSNCEGSVKV 227

Db 331 BEVILSPGHALSFIAAGEEHRHVCSNNFNLLSSRSHTIFTLMVESATGDEYD---GVIF 387

QY 228 SHMLVLAASERAAQTGAAGVRLKEGCNINRSFILGQVTKKLSGQGVGFNIRDSKL 287

Db 388 SQLMLIDLAGE--SSKTETTLGLRRKEGYSINKSLTLGTGVLKSEK-AATHIPYRDSKL 445

QY 288 TRILONSLGNPKTRITCTTPV--SFDETLTALQFASAKYMNTPYVNEVSTDEALLK 345

Db 446 TRLLQSSLSGHVSLCTITTPASSSEETHTLKFASRAKSIETIYASRNQIIDEKSLIK 505

QY 346 RYRKEIMDLKQLEEV-----SLETRAQAMEKQD--LAQLLEEKDLLOKQVNEK 392

Db 506 KYQREISTKLQLDQLRRGLVGVSHBELMSLKQLEEGGVKQMSRLEEEBAKALMSR 565

QY 393 IENUTRMLVTSS 404

Db 566 IQRLTKLILWST 577

RESULT 12

A54803

microtubule-associated motor KIF4 - mouse

N;Alternate names: kinesin-related protein KIF4

C;Species: Mus musculus (house mouse)

C;Date: 05-Apr-1995 #sequence_revision 05-Apr-1995 #text_change 02-Feb-2001

C;Accession: A54803; D44259

R;Sekine, Y.; Okada, Y.; Noda, Y.; Kondo, S.; Aizawa, H.; Takemura, R.; Hirokawa, N.

J. Cell Biol. 127, 187-201, 1994

A;Title: A novel microtubule-based motor protein (KIF4) for organelle transports, whose

A;Reference number: A54803; MUID:95014709; PMID:7929562

A;Accession: A54803

A;Molecule type: mRNA

A;Residues: 1-1231 <SEK>

A;Cross-references: GB:DI2646; NID:G563772; PIDN:BA002167.1; PID:dl002657; PID:G563773

R;Aizawa, H.; Sekine, Y.; Takemura, R.; Zhang, Z.; Nangaku, M.; Hirokawa, N.

J. Cell Biol. 119, 1287-1296, 1992

A;Title: Kinesin family in murine central nervous system.

A;Reference number: A44259; MUID:93077686; PMID:1447303

A;Accession: D44259

A;Status: nucleic acid sequence not shown; not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 91-111,'S',113-240 <AIZ>

A;Experimental source: brain

A;Note: sequence extracted from NCBI backbone (NCBIPI:118904)

C;Superfamily: unassigned kinesin-related proteins; kinesin motor domain homology

C;Keywords: ATP; nucleotide binding; P-loop

F;10-343/Domain: kinesin motor domain homology <RMOT>

F;88-95/Region: nucleotide-binding motif A (P-loop)

Query Match 33.8%; Score 693; DB 2; Length 1231;

Best Local Similarity 39.3%; Pred. No. 4.6e-37;

Matches 167; Conservative 69; Mismatches 143; Indels 46; Gaps 10;

QY 6 VAYCVVRPLNSREESLG-ETAQVYWKTDNNVYQVDSKSFNDFRVFHHGNETTKNVYEE 64

Db 10 VRVALRCPLVSKVEIKEGCQCLSFVPGEPQV--VGNDKSFYDFVDPSTQEVEVNT 67

QY 65 IAAPIDSAIQGYNGTIFAYGQTASGTYTMMGS-----EDH-----LGVIPRAIHDFQKIK 117

Db 68 AVAPLIKGVFKGNATVAYGQSGKTYSMGAYTAEOHSDSAIGVIRVLIQFLKEIN 127

QY 118 KPDPRELLRVSYMEIYNETITDLCGT-QKMKPLIREDVNRNVVADITEEVVYVSEM 176

Db 128 KKSDFEFTLVSVLEIYNEILDLSSREKATQINIREDPKGIKIVGLTEKTVLIVASD 187

QY 177 ALKMTWTKGSRHYGFTKMNQSRSHITIFRMILESREKGEPSNCGSVKVSHLNVLDA 236

Db 188 TVSCLEQGNNSRTVASTAMNSQSSRSHAIPTISIEQKK----NDRNSSFRSKHLVLDLA 243

QY 237 GSERAAQTGAAGVRLKEGCNINRSFILGQVTKKLSGQGVGFNIRDSKLTRILQNSLG 296

Db 244 GSERQKTKAEGDRLREGININRGLLCLGNVISALGDDKGNFVPRYRDSKLTLLQDSLG 303

QY 297 GNPKTRITCTTPV--SFDETLTALQFASAKYMNTPYVNEVSTDEALLKRYRKEIMDL 354

Db 304 GNSHTLMIACVSPADSNLEETLNTLYADRAKRIKPKPIIN-----IDPQAAELNHL 355

QY 355 KQLEEVSL-----ETRAQAMEKDQLAQLLEEKDLLOKQVNEKLEN 395

Db 356 KQVQVQIILLQAHGGTLPGDINVEPSENLQSLMEKNQ--SLVEENEKLSRGUSEAAGQ 413

QY 396 LTRML 400

Db 414 TAQWL 418

RESULT 13

H86350

hypothetical protein F8K7.17 - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001

C;Accession: H86350

R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso

ansen, N.F.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.

Nature 408, 816-820, 2000

A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, Z.A.; Luos, J.S.; Maiti, R.; Marziali,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A;Authors: Salzberg, S.B.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A;Reference number: A86141; MUID:21016719; PMID:11130712

A;Accession: H86350

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-909 <STO>

A;Cross-references: GB:AE005172; NID:G5263326; PIDN:AA41428.1; GSPDB:GN00141

C;Genetics:

A;Map position: 1

Query Match 33.7%; Score 692; DB 2; Length 909;

Best Local Similarity 39.3%; Pred. No. 3.5e-37;

Matches 167; Conservative 85; Mismatches 141; Indels 32; Gaps 10;

QY 6 VAYCVVRPLNSREESLG-ETAQVYWKTDNNVYQVDSKSFNDFRVFHHGNETTKNVYEEI 65

Db 76 ITVTIFRPLSPREVNGDEIAWYAGDVTIRNEYNSLCYGFDRVFGPTTTRVVYDIA 135

QY 66 AAPIDSAIQGYNGTIFAYGQTASGTYTMMGSDEHLGVIPRAIHDFQKIKF-----119

Db 136 AQOVVSGAMSGINGTVFAYGVTSSGKTHMHGQSPGIPLAVKDVFSIIQEVIANVVS 195

QY 120 ----PDRELLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNVVADITEEVVYVSE 175

Db 196 AIQTPEREFLRVSYLEIYNEIYNDLLDPTG--QNLRIRED--SQGTVEGKIDDEVLSPA 252

QY 176 MALKMTWTKG---KSRHYGETKMNQSRSHITIFRMILESREKGEPSNCGSVKVSHLNL 232

Db 253 HALSLIASGEVIAEHRHVGNNVNLFSRSHTMTITIESSPHGKGDGGE-DVLSQLHL 311

QY 233 VDLAGSERAQQTGAAGVRLKEGCNINRSFILGQVTKKLSGQGVGFNIRDSKLTRELQ 292

Db 312 IDLAGSE--SSKTEITQRRKEGSSINKSLTLGTVLSKLTDTK-AAHIFYRDSKLTRELQ 369

QY 293 NSLGNPKTRITCTTPVS--FDETLTALQFASAKYMNTPYVNEVSTDEALLKRYRKE 350

Db 370 STLSGHGRVSLCTITTPASSTSEETHNTLKAQRCHVEIKASRNKIMDEKSLIKRYQKE 429

QY 351 IMDLKKQLEEV-----SLETRAQAMEKDQLAQL-----LEEKDLLOKQVNEKLENITRM 399

Db 430 ISCLQBELTQLRHGNODDLADRLQVGLKSKCVKLSQSLRLEDEDEBAKALMGRIQLTKL 489

QY 400 LVTSS 404

Db 490 ILVST 494

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 09:26:36 ; Search time 3.88047 Seconds
(without alignments)
5421.082 Million cell updates/sec

Title: US-10-045-631B-88_COPY_2_405

Perfect score: 2053

Sequence: 1 AEGGAVVCVRVRPLNSREE.....LQKVQNEKIENLRMLVTSS 404

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2053	100.0	2663	1	CENE HUMAN	Q02224 homo sapien
2	723	35.2	701	1	KE3A_MOUSE	P28741 mus musculus
3	716.5	34.9	702	1	KE3A_HUMAN	Q9Y496 homo sapien
4	714	34.8	747	1	KE3B_HUMAN	O15066 homo sapien
5	713	34.7	747	1	KE3B_MOUSE	O61771 mus musculus
6	712.5	34.7	742	1	KL21_STRPU	P46871 strongyloce
7	708.5	34.5	785	1	FL10_CHLRE	P46869 chlamydomon
8	702	34.2	699	1	KL22_STRPU	P46872 strongyloce
9	694.5	33.8	1232	1	KE4A_HUMAN	O95239 homo sapien
10	693	33.8	1231	1	KE4A_MOUSE	P33174 mus musculus
11	691.5	33.7	1029	1	KE17_HUMAN	Q9p2e2 homo sapien
12	687	33.5	1226	1	KE4A_XENLA	Q31784 xenopus lae
13	683	33.3	1038	1	KE17_MOUSE	Q99pw8 mus musculus
14	667	32.5	1225	1	KE4A_CHICK	Q90640 gallus gall
15	662.5	32.3	1066	1	KL61_DROME	P46863 drosophila
16	656	32.0	672	1	OSM3_CAEEL	P46873 caenorhabdi
17	652	31.8	975	1	KINH_DROME	P17210 drosophila
18	651	31.7	796	1	KE3C_MOUSE	O35066 mus musculus
19	649.5	31.6	1749	1	KL3A_MOUSE	Q9eqw7 mus musculus
20	649	31.6	796	1	KE3C_RAT	O55165 rattus norv
21	648.5	31.6	793	1	KE3C_HUMAN	O14782 homo sapien
22	645	31.4	1031	1	KINH_STRPU	P35978 strongyloce
23	644.5	31.4	1805	1	KL3A_HUMAN	O9h1h9 homo sapien
24	643.5	31.3	1826	1	KL3B_HUMAN	O9ngt8 homo sapien
25	637	31.0	967	1	KINH_LOLPE	P21613 loligo peal
26	634.5	30.9	1032	1	KINH_HUMAN	O12840 homo sapien
27	631.5	30.8	957	1	KE5C_HUMAN	O60282 homo sapien
28	631.5	30.8	1027	1	KINH_MOUSE	P33175 mus musculus
29	630.5	30.7	956	1	KE5C_MOUSE	P28738 mus musculus
30	630.5	30.7	963	1	KINH_MOUSE	O61768 mus musculus
31	629.5	30.7	963	1	KINH_HUMAN	P33176 homo sapien
32	625	30.4	784	1	KL68_DROME	P46867 drosophila
33	622	30.3	815	1	KINH_CAEEL	P34540 caenorhabdi

ALIGNMENTS

RESULT 1

ID	CENE_HUMAN	STANDARD;	PRT;	2663 AA.
AC	Q02224;			
DT	01-JUL-1993	(Rel. 26, Created)		
DT	01-JUL-1993	(Rel. 26, Last sequence update)		
DT	10-OCT-2003	(Rel. 42, Last annotation update)		
DE	Centromeric protein B (CENP-E protein).			
GN	CENPE.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
EX	MEDLINE=93024922; PubMed=1406971;			
RA	Yen T.J.; Li G.; Schaar B.T.; Szilak I.; Cleveland D.W.;			
RT	"CENP-E is a putative kinetochore motor that accumulates just before			
RT	mitosis.";			
RL	Nature 359:536-539 (1992).			
RN	[2]			
RP	CHARACTERIZATION.			
RX	MEDLINE=95196755; PubMed=7889940;			
RA	Thrower D.A.; Jordan M.A.; Schaar B.T.; Yen T.J.; Wilson L.;			
RT	"Mitotic HeLa cells contain a CENP-E-associated minus end-directed			
RL	microtubule motor.";			
RN	EMBO J. 14:918-926 (1995).			
RP	[3]			
RX	CHARACTERIZATION.			
RA	MEDLINE=98437347; PubMed=9763420;			
RA	Chan G.K.T.; Schaar B.T.; Yen T.J.;			
RT	"Characterization of the kinetochore binding domain of CENP-E reveals			
EL	interactions with the kinetochore proteins CENP-F and hBUBR1.";			
RN	J. Cell Biol. 143:49-63 (1998).			
RP	[4]			
RX	FARNESYLATION.			
RA	MEDLINE=20459117; PubMed=10852915;			
RA	Ashar H.R.; James L.; Gray K.; Carr D.; Armstrong L.;			
RT	Bishop W.R.; Kirschmeier P.;			
RT	"Farnesyl transferase inhibitors block the farnesylation of CENP-E			
RT	and CENP-F and alter the association of CENP-E with the			
RT	microtubules.";			
RL	J. Biol. Chem. 275:30451-30457 (2000).			
CC	-!- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE			
CC	KINETOCORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE			
CC	OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT			
CC	AND/OR SPINDLE ELONGATION.			
CC	-!- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.			
CC	-!- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCORES DURING			
CC	CONGRESSION. RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS			
CC	QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.			
CC	-!- SIMILARITY: Belongs to the kinesin-like protein family.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			

P23678 caenorhabdi
Q9us60 schizosacch
P48467 neurospora
P28025 xenopus lae
Q91783 xenopus lae
P52732 homo sapien
O43093 syncephalas
Q15058 homo sapien
P28742 saccharomyc
P82266 arabidopsis
Q12756 homo sapien
P17120 emericella

34 622 30.3 1584 1 U104_CAEEL
35 620 30.2 554 1 KLP3_SCHPO
36 609.5 29.7 928 1 KINH_NEUCR
37 609.5 29.7 1060 1 EGS1_XENLA
38 609 29.7 1067 1 EG52_XENLA
39 606.5 29.5 1057 1 KF11_HUMAN
40 603 29.4 935 1 KINH_SYNRA
41 602.5 29.3 1648 1 KF14_HUMAN
42 601.5 29.3 1111 1 KIP1_YEAST
43 596 29.0 1056 1 KF15_ARATH
44 591.5 28.8 1690 1 KF1A_HUMAN
45 591 28.8 1184 1 BIMC_EWENI

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; Z15005; CAA78727.1; -;
PIR; S28261; S28261.
HSP; P17119; 3KAR.
Genew; HGNC:1856; CENPE.
GK; Q02224; -;
MIM; 117143; -;
GO; GO:0005699; C:kinetochore; TAS.
GO; GO:0005634; C:nucleus; TAS.
GO; GO:0008350; F:kinetochore motor activity; TAS.
GO; GO:0000067; P:DNA replication and chromosome cycle; TAS.
GO; GO:0007079; P:mitotic chromosome movement; TAS.
GO; GO:0007080; P:mitotic metaphase plate congression; TAS.
InterPro; IPR001752; kinesin_motor.
Pfam; PF00225; kinesin; 1.
PRINTS; PR00380; KINESINHEAVY.
SMART; SM00129; KISC; 1.
PROSITE; PS00411; KINESIN MOTOR DOMAIN; 1.
PROSITE; PS00067; KINESIN MOTOR DOMAIN; 1.
Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;
Cell cycle; Centromere; Lipoprotein; Prenylation.
DOMAIN 1 335 KINESIN-MOTOR.
DOMAIN 336 2471 COILED COIL (POTENTIAL).
DOMAIN 2472 2663 GLOBULAR (POTENTIAL).
NP BIND 86 93 ATP (BY SIMILARITY).
LIPID 2660 2660 S-farnesyl cysteine.
SEQUENCE 2663 AA; 312087 MW; CEFC13880C8C8B8 CRC64;

Query Match 100.0%; Score 2053; DB 1; Length 2663;
Best Local Similarity 100.0%; Pred. No. 3.6e-123;
Matches 404; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AEEGAVAVCVVRPLNSREESLGETAQVYKTDNNVYQVDSKSFNDFRVFHNQETTKN 60
Db 2 AEEGAVAVCVVRPLNSREESLGETAQVYKTDNNVYQVDSKSFNDFRVFHNQETTKN 61
QY 61 VYBEIAAPIIDSAIQGYNGTIFAYGQTASGTYTMMGSEDLGVIPRAIHDIQKIKKFP 120
Db 62 VYBEIAAPIIDSAIQGYNGTIFAYGQTASGTYTMMGSEDLGVIPRAIHDIQKIKKFP 121
QY 121 DREFLLRVSYMEIYNETITDLCTQKMKPLIREDVNRNRYVADLVEEVVYTSMAK 180
Db 122 DREFLLRVSYMEIYNETITDLCTQKMKPLIREDVNRNRYVADLVEEVVYTSMAK 181
QY 181 ITKGEKSRHYGETKQNRSSRSHITFRMILESRKGEPSNCEGSKVSHLNLVDLAGSER 240
Db 182 ITKGEKSRHYGETKQNRSSRSHITFRMILESRKGEPSNCEGSKVSHLNLVDLAGSER 241
QY 241 AAOQTGAAGVRLKEGCNINRSFLIGQVTKLSDGVGFIYRDSKLTIRLQNSLGNPK 300
Db 242 AAOQTGAAGVRLKEGCNINRSFLIGQVTKLSDGVGFIYRDSKLTIRLQNSLGNPK 301
QY 301 TRIICITTPVSFDTLQAFSTAKYKNTPYNEVSTDEALLKRYRKEIMDLKQLEE 360
Db 302 TRIICITTPVSFDTLQAFSTAKYKNTPYNEVSTDEALLKRYRKEIMDLKQLEE 361
QY 361 VSLTEAQAAMEKDQALKEELKLLQKQVNEKIENLRLMTVSS 404
Db 362 VSLTEAQAAMEKDQALKEELKLLQKQVNEKIENLRLMTVSS 405

RESULT 2
KF3A_MOUSE
ID KF3A_MOUSE STANDARD; PRT; 701 AA.
AC P28741;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Kinesin-like protein KIF3A (Microtubule plus end-directed kinesin

DE motor 3A).
GN KIF3A OR KIF3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A.
RP TISSUE=Brain;
RC MEDLINE=93077686; PubMed=1447303;
RA Aizawa H., Sekine Y., Takemura R., Zhang Z., Nangaku M., Hirokawa N.;
RT "Kinesin family in murine central nervous system.";
RL J. Cell Biol. 119:1287-1296 (1992).
CC -!- FUNCTION: MICROTUBULE-BASED ANTEROGRADE TRANSLOCATOR FOR
MEMBRANOUS ORGANELLES. PLUS END-DIRECTED MICROTUBULE SLIDING
ACTIVITY IN VITRO.
CC -!- SUBUNIT: HETERODIMER OF KIF3A AND KIF3B.
CC -!- TISSUE SPECIFICITY: EXPRESSED ALMOST EXCLUSIVELY IN ADULT BRAIN
TISSUE (MAINLY IN THE CEREBELLAR GRANULAR LAYER) WITHIN A SINGLE
TYPE OF NEURONAL CELL.
CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
II SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; D12645; BAA02166.1; -;
PIR; B44259; B44259.
HSP; P17119; 3KAR.
MGD; MGI:107689; Kif3a.
InterPro; IPR001752; Kinesin_motor.
Pfam; PF00225; kinesin; 1.
PRINTS; PR00380; KINESINHEAVY.
SMART; SM00129; KISC; 1.
PROSITE; PS00411; KINESIN MOTOR DOMAIN; 1.
PROSITE; PS00067; KINESIN MOTOR DOMAIN; 1.
Motor protein; Microtubule; ATP-binding; Coiled coil; Neurone.
KINESIN-MOTOR (BY SIMILARITY).
DOMAIN 1 350 COILED COIL (BY SIMILARITY).
DOMAIN 351 598 GLOBULAR.
DOMAIN 599 701 ATP (BY SIMILARITY).
NP BIND 100 107 POLY-ARG.
DOMAIN 442 445 POLY-GLU.
DOMAIN 509 512 POLY-ARG.
SEQUENCE 701 AA; 80167 MW; 2405872DF2D85A29 CRC64;
Query Match 35.2%; Score 723; DB 1; Length 701;
Best Local Similarity 44.4%; Pred. No. 6.7e-39;
Matches 178; Conservative 66; Mismatches 137; Indels 20; Gaps 10;
QY 6 VAVCVVRPLNSREESL--GETAQVYKTDNNVYQVDSG----KSFNDFRVFHNQETTK 59
Db 15 KVVRVCRPLNREKSKCYQAVSDVEMRGITVHKTDSSNEPKTFTFDVFPESKQL 74
QY 60 NVYBEIAAPIIDSAIQGYNGTIFAYGQTASGTYTMMGSE---DHLGVIPRAIHDIQKI 116
Db 75 DVYNLTARPIIDSLVLEGVNGTIFAYGQTGKTGTMEGVRAVPLGLRGVIFNSFAHIFGHI 134
QY 117 KFFP--DREFLLRVSYMEIYNETITDLCTQKMKPLIREDVNRNRYVADLVEEVVYTS 175
Db 135 AKAEGRDTRFLVRVSYLEIYNEEVREDLL--GKQDQORLEKRPDPGVYKILSATVNNAD 193
QY 176 MALKWIITKGEKSRHYGETKQNRSSRSHITFRMILESRKGEPSNCEGSKVSHLNLVDL 235
Db 194 DMDRIMTLGHKVRSGVATNNMHSRSHAIITITIECKEKGVDGNM--VRMGKHLVDL 251
QY 236 AGSERRAQTGAAGVRLKEGCNINRSFLIGQVTKLSDGVGFIYRDSKLTIRLQNSL 295
Db 252 AGSERQANTGATGQRLKEATKINLSLSTGLNVSALVDGK--STHVPYRNSKLTIRLQNSL 310


```

DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE Kinesin-like protein KIF3B (Microtubule plus end-directed kinesin
DE motor 3B) (HH0048).
GN KIF3B OR KIAA0359.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ON NCBI_TaxID=9606;
RX [1]
RX SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=97349984; PubMed=9205841;
RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 4:141-150(1997).
[2]
RX SEQUENCE FROM N.A.
RX MEDLINE=21638749; PubMed=11780052;
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavros G., Almeida J.P., Babbage A.K., Bagguley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P., Clee C.M.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA Hammond S., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Huckle E., Hurler J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA LeHaevela M.H., Levensha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConnell L.J., McLay K., McMurray A.A.,
RA Milne S., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehara H.K., Showkneen R., Sims S.,
RA Stuke C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.;
RT "The DNA sequence and comparative analysis of human chromosome 20.";
RL Nature 414:865-871(2001).
[3]
RX IDENTIFICATION IN A COMPLEX WITH SMC3 AND KIFAP3B.
RX MEDLINE=98175913; PubMed=9506951;
RX Shimizu K., Shiratani H., Honda T., Minami S., Takai Y.;
RT "Complex formation of SNAP/KAP3, a KIF3A/B ATPase motor-associated
RT protein, with a human chromosome-associated polypeptide.";
RL J. Biol. Chem. 273:6591-6594(1998).
CC -!- FUNCTION: Involved in tethering the chromosomes to the spindle
CC pole and in chromosome movement. Microtubule-based anterograde
CC translocator for membranous organelles. Plus end-directed
CC microtubule sliding activity in vitro (By similarity).
CC -!- SUBUNIT: Heterodimer of KIF3A and KIF3B (By similarity). Interacts
CC with the SMC3 subunit of the cohesin complex.
CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
CC II SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC

```

```

CC -----BAA20815.1; --
DR EMBL; AB002357; BAA20815.1; --
DR EMBL; AL121897; CAC16425.1; --
DR HSP; P17119; 3KAR.
DR Genew; HGNC:6320; KIF3B.
DR MIM; 603754; --
DR GO; GO:0005873; C:plus-end kinesin complex; TAS.
DR GO; GO:0003777; F:Microtubule motor activity; TAS.
DR GO; GO:0008574; F:plus-end-directed kinesin ATPase activity; TAS.
DR GO; GO:0008089; P:anterograde axon cargo transport; TAS.
DR GO; GO:0007368; P:determination of left/right asymmetry; TAS.
DR InterPro; IPR001752; kinesin_motor.
DR Pfam; PF00225; kinesin; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
DR PROSITE; PS00667; KINESIN MOTOR DOMAIN2; 1.
KW Motor protein; Microtubule; ATP-binding; Coiled coil; Neurone.
FT DOMAIN 1 345 KINESIN-MOTOR (BY SIMILARITY).
FT DOMAIN 346 579 COILED COIL.
FT DOMAIN 580 747 GLOBULAR.
FT NP BIND 96 103 ATP (POTENTIAL).
FT DOMAIN 386 393 POLY-GLY.
FT DOMAIN 394 406 POLY-GLU.
FT DOMAIN 723 730 POLY-SER.
FT SEQUENCE 747 AA; 85125 MW; 97FA4573AFA87023 CRC64;
SQ
Query Match 34.8%; Score 714; DB 1; Length 747;
Best Local Similarity 41.4%; Pred. No. 2.7e-38;
Matches 177; Conservative 69; Mismatches 144; Indels 38; Gaps 10;
QY 5 AVAVCVVRPLNRSRESLG-----ETAVYVTKDNNVIYQVDSKSPNDRVPHGN 55
DB 9 SVRVVRCRPMNGEKAASYDKVDVVDVGLGVSVKPNKGTATHEM--PKTFTFDVVDWN 66
QY 56 ETTKNVYEETAAPIIDSAIOGYNGTTFAYGQTASGKTYTM---MGSEDLHLGVIPRAIHDI 112
DB 67 AKOFELYDETRFPLNDSVLQGFNGTTFAYGQTGKTYTMGIRGDPEKRGVIPSFDHI 126
QY 113 FQKIKKPPDFEFLIRSVMEIYNETITDLCGQTKMKPLIREDVNRNVVADLTEVVY 172
DB 127 FTHLSRSONQQLVRSAYLIEYBEIRDLLSKDQ--TKRLKELKRPDTGVYVKDLSFFVTK 185
QY 173 TSEALKWIKYGEKSRHYGTAKNORSRSHYTFRMILESEKGEPCSVKSHNL 232
DB 186 SVKEIEHVMVGNQNSVGATNMNHSRSHAFVITIECEVG--LDGENHVRGKLN 243
QY 233 VDLAGSRAAQTGAAGVRLKEGCMNRSFLTGQVKKLSDGVGGGFNYRDSKLRLQ 292
DB 244 VDLAGSERQAKTGAQGERLKEATKINLSLGNVSNALVDGK--STHPIYRDSKLRLQ 302
QY 293 NSLGGNPKTRICTITPVSF--DETLALQOPASTAKYMKNTPPYNEVSTDEALLKRYKE 350
DB 303 DSLGNAKTVMVANVPASYNVEETLTLRYANRAKAIKKNKPVNEDPKD--ALLREFOEE 361
QY 351 IMDLKKQLSEVSLSTRAQ-----AMEKDLQALLEE---KDLQKQVNEKI 393
DB 362 IARLKAQLKRSIGRRKRKRREGGGGGGGGEGEGEGEGEGEGEGEGEGEGEGEGEG 421
QY 394 ENLTRLV 401
DB 422 EIEKRAIV 429
RESULT 5
KF3B_MOUSE
ID KF3B_MOUSE STANDARD; PRT; 747 AA.
AC Q61771;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Kinesin-like protein KIF3B (Microtubule plus end-directed kinesin
DE motor 3B).

```

GN KIF3B.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ICR; TISSUE=Brain;
 RX MEDLINE=96032268; PubMed=7559760;
 RA Yamazaki H., Nakata T., Okada Y., Hirokawa N.;
 RT "KIF3A/B: a heterodimeric kinesin superfamily protein that works as a
 RT microtubule plus end-directed motor for membrane organelle
 RT transport."; J. Cell Biol. 130:1387-1399(1995).
 RL J. Cell Biol. 130:1387-1399(1995).
 CC -!- FUNCTION: Involved in tethering the chromosomes to the spindle
 CC pole and in chromosome movement. Microtubule-based anterograde
 CC translocator for membranous organelles. Plus end-directed
 CC microtubule sliding activity in vitro (By similarity).
 CC -!- SUBUNIT: Interacts with the SMC3 subunit of the cohesin
 CC complex (By similarity). Heterodimer of KIF3A and KIF3B.
 CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
 CC II SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; D26077; BAA05070.1; -
 DR PIR; A57107; A57107.
 DR HSSP; P17119; 3KAR.
 DR MGD; MGI:107688; Kif3b.
 DR InterPro; IPR001752; kinesin_motor.
 DR Pfam; PF00225; kinesin; 1.
 DR PRINTS; PR00380; KINSEINHEAVY.
 DR SMART; SM00129; KISC; 1.
 DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
 DR PROSITE; PS0067; KINESIN MOTOR DOMAIN2; 1.
 KW Motor protein; Microtubule; ATP-binding; Coiled coil; Neurone.
 FT DOMAIN 1 345 KINESIN-MOTOR (BY SIMILARITY).
 FT COILED COIL (BY SIMILARITY).
 FT GLOBULAR.
 FT NP BIND 96 103 ATP (POTENTIAL).
 FT DOMAIN 386 393 POLY-GLY.
 FT DOMAIN 394 405 POLY-GLU.
 FT DOMAIN 723 730 POLY-SER.
 SQ SEQUENCE 747 AA; 85288 MW; FA369A4190EC8B47 CRC64;
 Query Match 34.7%; Score 713; DB 1; Length 747;
 Best Local Similarity 41.4%; Pred. No. 3.2e-38;
 Matches 177; Conservative 69; Mismatches 144; Indels 38; Gaps 10;
 QY 5 AVAVCVVRPLNSREESLG-----ETAQVYWKTDNNVIYQVDSKGFNDFRVHGN 55
 DB 9 SVRVVVRCPWNGKEKAASQVVDVVKLGQVSKNPKGTSHEM--PKTFTFVAVDWN 66
 QY 56 ETTKNVVEETAAPIIDSAIOGYNGTIFAYGQTASGKTYTM---MGSEDHGLVIPRAIHI 112
 DB 67 AKOFELYDETFRPLVDVSLQGFNGTIFAYGQTGRTGTYTMEGVGRDPEKRGVIPSFDHI 126
 QY 113 FQRIKPPDFREFLRVSMYELNETITDLCCGTQMKPLIREDVNVNVDLTSEVYV 172
 DB 127 FTHLSRQOQLVRSYLIYBEIRDLSSKQ--TKRLKRPOTGVTVKOLSSFEVK 185
 QY 173 TSEWALKWITKESRHYGTHKQNRSSRSHITFRMILESREKGEPCNSGVKVSHLN 232
 DB 186 SVKEIEHVMVNGQNRSGVATNNMHSRSHAFVITIECEVG--LDGENHVRVGLKL 243
 QY 233 VDLAGSERAQCTAAGVRLKGGCNINSLFILGQVKKLSGQGVGFINTRDSKLTRELQ 292

Db 244 VDLAGSERAQCTAAGQERLKEATKINLSALGNVISALVDGK-STHPIYRDSKLTRELQ 302
 QY 293 NSLGGNPKTRIICTIPVVF--DETALQAFSTAKYMKNTPYVNEVSTDEALLKRYRKE 350
 DB 303 DSLGGNAKTMVANVGVPASVYVEETLTLLRYANRAKNKKNPRVNDPKD-ALLREFOEE 361
 QY 351 IMDLKKOLEVSLFETFAQ-----AMEKDQLAQLLEE-----KOLLOKVONEKI 393
 DB 362 IARLKAQLEKRIIGRRKRREKRREGSGGGGEGEEGEGEDGDKDDYVREQQEK 421
 QY 394 ENLTRLIV 401
 DB 422 ETEKRAIV 429
 RESULT 6
 ID K121 STRPU STANDARD; PRT; 742 AA.
 AC P46871;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Kinesin-II 95 kDa subunit (KRP-85/95 95 kDa subunit).
 GN KRP95.
 OS Strongylocentrotus purpuratus (Purple sea urchin).
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
 OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
 OC Strongylocentrotus.
 OX NCBI_TaxID=7668;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Egg;
 RX MEDLINE=94050179; PubMed=8232586;
 RA Cole D.G., Chinn S.W., Wedaman K.P., Hall K., Vuong T.,
 RA Scholey J.M.;
 RT "Novel heterotrimeric kinesin-related protein purified from sea
 RT urchin eggs";
 RL Nature 366:268-270(1993).
 CC -!- SUBUNIT: Heterotrimer of a 115 kDa subunit (KAP115) and two
 CC kinesin-like subunits of 95 kDa (KRP95) and 85 kDa (KRP85).
 CC -!- PTM: The N-terminus is blocked.
 CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
 CC II SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; U00996; AAA87393.1; -
 DR HSSP; P17119; 3KAR.
 DR InterPro; IPR001752; kinesin_motor.
 DR Pfam; PF00225; kinesin; 1.
 DR PRINTS; PR00380; KINSEINHEAVY.
 DR SMART; SM00129; KISC; 1.
 DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
 DR PROSITE; PS0067; KINESIN MOTOR DOMAIN2; 1.
 KW Motor protein; Microtubule; ATP-binding; Coiled coil.
 FT DOMAIN 1 337 KINESIN-MOTOR (BY SIMILARITY).
 FT DOMAIN 338 613 COILED COIL (BY SIMILARITY).
 FT DOMAIN 614 742 GLOBULAR (BY SIMILARITY).
 FT NP BIND 95 102 ATP (POTENTIAL).
 SQ SEQUENCE 742 AA; 84202 MW; 47C40A367BAA77B5 CRC64;
 Query Match 34.7%; Score 712.5; DB 1; Length 742;
 Best Local Similarity 41.3%; Pred. No. 3.4e-38;
 Matches 175; Conservative 65; Mismatches 137; Indels 47; Gaps 10;
 QY 6 VAVCVVRPLNSREESLGETAQVYWKTDNNVIYQV-----DGSKSFNDFRVHGN 58

Db 9 VKVVRCPMNSKETSCQHKRIVEMDNKRGLVEVTPKPPGEPNKSFTFDVYDWSKQ 68
 QY 59 KNYVEIAPITDSALQYNGTIFAYGOTASCTYTMG--SEDLGVIPRAIHDFQK 115
 Db 69 IDLYDETFRSLVESVLOQNGTIFAYGOTGKTFTMEGVRSNPELRGVIPNPFHIFH 128
 QY 116 IKKFPDRFLLVSVMEIYNETITDLCCTQKMKPLIREDVNRNYYVADLTFEEVYTS 175
 Db 129 IARTNQQLFVRSVLEIYQEBIRDLAKDQK-KRLDKERPDPTGVYVVDLSFVTKSVK 187
 QY 176 MALKWITKGEKSRHYGETKQNRSSRSHTIFRMILESREKGRPSNCEGSKVYSHNLVDL 235
 Db 188 EIEHVMTVGNNSRVSGTNNBHSRSHAFTITIECSLGVDG--ENHIRVGKLNVDL 245
 QY 236 AGSERAQACAGVRLKEGCNINRSFILGOVTKLSDGQVGFNVRDSKLTIRIQLNSL 295
 Db 246 AGSERAQAKTGATGDRLEKATKINLSALGNVISALVDGK-SSHIPYRDSKLTIRIQLNSL 304
 QY 296 GGNPKTRITCTTPVS--FDETLTALQFASTAKYMKNTPYVNEVSTDEALLKRYRKEIMD 353
 Db 305 GGNKATVWVNMGPASVNFDETITTLRYANRAKNIKNPKINEDPKD-ALLREFQBEISR 363
 QY 354 LKKQL-----EVSLTRAQAMEKDQLAQLEE--KDLLOKV 388
 Db 364 LKQALDKGPGDGRKKGKRRKPGEGQGGDDIDIEETEEGEMDE-----EETKESQOQKL 418
 QY 389 QNEK 392
 Db 419 EEEK 422

RESULT 7

FL10_CHRE STANDARD; PRT; 786 AA.
 AC P46869;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Kinesin-like protein FLA10 (KHP1 protein).
 GN FLA10.
 OS Chlamydomonas reinhardtii.
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 OC Chlamydomonadaceae; Chlamydomonas.
 OX NCBI_TaxID=3055;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=137;
 RA MEDLINE=94299638; PubMed=8027176;
 RX Walther Z., Vashishtha M., Hall J.L.;
 RT "The Chlamydomonas FLA10 gene encodes a novel kinesin-homologous protein.";
 RL J. Cell Biol. 126:175-188(1994).
 CC -!- FUNCTION: Probably involved in flagellar assembly and maintenance.
 CC May play a role in flagellar synthesis.
 CC -!- TISSUE SPECIFICITY: Flagellar axoneme.
 CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
 CC II SUBFAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; L33697; AAA21738.1; -
 DR PIR; A53939; A53939.
 DR HSP; P17119; 3KAR.
 DR InterPro; IPR001752; kinesin_motor.
 DR Pfam; PF00225; kinesin; 1.
 DR PRINTS; PR00380; KINESINHEAVY.
 DR SMART; SM00129; KISC; 1.

DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
 DR PROSITE; PS0067; KINESIN_MOTOR_DOMAIN2; 1.
 KW Motor protein; Microtubule; ATP-binding; Coiled coil.
 FT DOMAIN 1 358 KINESIN_MOTOR (BY SIMILARITY).
 FT DOMAIN 367 687 COILED COIL (POTENTIAL).
 FT DOMAIN 688 766 GLOBULAR (POTENTIAL).
 FT NP_BIND 97 104 ATP (POTENTIAL).
 FT DOMAIN 388 391 POLY-GLY.
 FT DOMAIN 705 714 POLY-GLY.
 FT DOMAIN 756 759 POLY-ASP.
 SQ SEQUENCE 786 AA; 86671 MW; F90969203EB79F1B CRC64;

Query Match 34.5%; Score 708.5; DB 1; Length 786;
 Best Local Similarity 44.4%; Pred. No. 6.6e-38;
 Matches 167; Conservative 66; Mismatches 118; Indels 25; Gaps 9;

QY 5 AVAVCVVRPLNRSRESLGETAQVWKTNVY-----QVDGS---KSNFDRVFHNET 57
 Db 10 SVKVVVRCPLNGKEKADGRSRIVDMVDAGQVKVNPKNKADASEPPKAFTFDQVYDWSKQ 69
 QY 58 TKNVYEEIAPITDSALQYNGTIFAYGOTASCTYTMGSEDH--LGVIPRAIHDFQ 114
 Db 70 QRDVFDITARPLDSCIEGNGTIFAYGOTGKTFTMEGVRSNPELRGVIPNPFHIFH 129
 QY 115 KI-KKFPDRFLLVSVMEIYNETITDLCCTQKMKPLIREDVNRNYYVADLTFEEVYTS 173
 Db 130 IARDNQQLFVRSVLEIYQEBIRDLAKDQK-KRLDKERPDPTGVYVVDLSFVTKSVK 188
 QY 174 SEMALKWITKGEKSRHYGETKQNRSSRSHTIFRMILESREK-----GEPSNCEGS 224
 Db 189 YEEMNKVLLAGKDNQVGTATLNQSSRSHTITIECTIEKLESAAAQPGAKKDDSNH 248
 QY 225 VKVSHNLVDLAGSERAQACAGVRLKEGCNINRSFILGOVTKLSDGQVGFNVRDSKLTIRIQLNSL 284
 Db 249 VRVSKLNVDLAGSERQDKTGATGDRLEKATKINLSALGNVISALVDGK-SSHIPYRDSKLTIRIQLNSL 307
 QY 285 SKLTIRIQLNSLGNPKTRITCTTPV--SFDETLTALQFASTAKYMKNTPYVNEVSTDEA 342
 Db 308 SKLTIRIQLNSLGNPKTRITCTTPV--SFDETLTALQFASTAKYMKNTPYVNEVSTDEA 366
 QY 343 LLKRYRKEIMDLKKQL 358
 Db 367 MLRFQFEIEIKKLEQL 382

RESULT 8

KI22_STRPU STANDARD; PRT; 699 AA.
 AC P46872;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Kinesin-II 85 kDa subunit (KRP-85/95 85 kDa subunit).
 GN KRP85.
 OS Strongylocentrotus purpuratus (Purple sea urchin).
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
 OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
 OC Strongylocentrotus.
 OX NCBI_TaxID=7668;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Egg;
 RX MEDLINE=94050179; PubMed=8232586;
 RA Cole D.G., Chinn S.W., Wedaman K.P., Hall K., Vuong T.,
 RA Scholey J.M.;
 RT "Novel heterotrimeric kinesin-related protein purified from sea urchin eggs."
 RL Nature 366:268-270(1993).
 CC -!- SUBUNIT: Heterotrimer of a 115 kDa subunit (KAP115) and two
 CC kinesin-like subunits of 95 kDa (KRP95) and 85 kDa (KRP85).
 CC -!- PTM: The N-terminus is blocked.
 CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
 CC II SUBFAMILY.

```

-----
CC CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR DR EMBL; LI6993; AAA16098.1; -.
DR DR PIR; S38982; S38982.
DR DR HSSP; P17119; 3KAR.
DR DR InterPro; IPR001752; kinesin_motor.
DR DR Pfam; PF00225; kinesin; 1.
DR DR PRINTS; PRO0380; KINESINHEAVY.
DR DR SMART; SM00129; KISC; 1.
DR DR PROSITE; PS00411; KINESIN MOTOR DOMAIN; 1.
DR DR PROSITE; PS50067; KINESIN MOTOR DOMAIN; 1.
DR DR Motor protein; Microtubule; ATP-binding; Coiled coil.
FT DOMAIN 1 340 KINESIN-MOTOR (BY SIMILARITY).
FT DOMAIN 341 619 COILED COIL (BY SIMILARITY).
FT DOMAIN 620 699 GLOBULAR (BY SIMILARITY).
FT NP BIND 97 104 ATP (POTENTIAL).
SQ SEQUENCE 699 AA; 78697 MW; 7B3966111C808190 CRC64;

Query Match 34.2%; Score 702; DB 1; Length 699;
Best Local Similarity 39.7%; Pred. No. 1.5e-37;
Matches 190; Conservative 65; Mismatches 131; Indels 92; Gaps 16;

Qy 6 VAVCVVRPLNRSSEISG-----ETAOVYWKTDNNVIYQVDSKSNFNRVPHG 54
Db 11 VRVVRCPNLNSETGGFKSVVMDMGTGVQV---TNPNA-PGSEPPKSFTEFTVAP 66

Qy 55 NETKNVYEEIAAPIIDSAIQGNGTIFAYGOTASGKTYTMWG--SEDLH-GVPIRAIHD 111
Db 67 GAKQTVDYNTARPIVDALIEGNGTIFAYGQTGKTFTMEGVRSQPELRGIIINPSFAH 126

Qy 112 IFQIKIKFPDR-EFLRVSYMEIYNTITDLCGTQOMKPLIIREDVNRNVYVADLTVEV 170
Db 127 IFGHIAKEQENRVFLRVYSLEYTNEVEVDLL-GKDDQQRHLEVKRPPDVGVYVKDLSAFV 185

Qy 171 VYTESMALAKWTKGEKSRHYGETKMNORSRSHTFPMILLESKEGPESGNCEGVKVSHL 230
Db 186 VNNADDMDRIMTLGNKRSVGNATMNSSSRSHAFTTLERSDMG--LDKEQHVVRGKL 243

Qy 231 NLVDLAGSRAAQTAGVRLKEGNCNINRSLPILGQVKKLSDGQGVGFVNRDSKLTPI 290
Db 244 HWVDLAGSRQTKATGQRLKEATKINLSLTGNVSSLDVGK-STHPIYRNSKLTFL 302

Qy 291 LQNSLGNPKTRIICTITPV--SFDETLTALQFATKMKNTPIVNEVSTDEALLKRYR 348
Db 303 LQDSLGNNAKTVNCANIGPAEYNYDETISTURYANRAKNIKAKINEDPDK-ALLREFQ 361

Qy 349 KEIMDLKKQLEE-----VSLSTRQAM 370
Db 362 KEIEELKKQISSESGRLDDDESGSESGDEAGEGVKKRKGKPKRKLSPIMIA-AM 420

Qy 371 EK--DQLAQLLBE-KDLL-----QKVONEKIENLARMLEV 401
Db 421 QKKIDEEKKALEEKDMVEEDRNTVHRELQRRESELHKAQDDQKILNEKLNAIQKKLI 478

RESULT 9
KF4A HUMAN
ID KF4A HUMAN STANDARD; PRT; 1232 AA.
AC O95239; Q9NNY6; Q9NY24; Q9UWW3;
DT 15-JUL-1999 (Rel. 38, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chromosome-associated kinesin KIF4A (Chromokinesin).
GN KIF4A OR KIF4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

FT CONFLICT 564 564 L -> P (IN REF. 2).
 FT CONFLICT 600 600 K -> E (IN REF. 3).
 FT CONFLICT 668 668 R -> K (IN REF. 3 AND 4).
 FT CONFLICT 928 928 Q -> P (IN REF. 1).
 FT CONFLICT 958 958 Q -> R (IN REF. 3).
 FT CONFLICT 960 960 L -> Q (IN REF. 1).
 FT CONFLICT 996 997 LL -> S (IN REF. 4).
 FT CONFLICT 1003 1014 OKHLPKDTLSP -> RTLPRIPFYLQ (IN REF. 4).
 FT CONFLICT 1022 1022 P -> Q (IN REF. 2).
 FT CONFLICT 1077 1077 K -> N (IN REF. 2).
 FT CONFLICT 1138 1138 G -> S (IN REF. 2).
 SQ SEQUENCE 1232 AA; 139908 MW; FF74052A17A8E8F7 CRC64;

Query Match 33.8%; Score 694.5; DB 1; Length 1232;
 Best Local Similarity 39.8%; Pred. No. 9.3e-37;
 Matches 168; Conservative 66; Mismatches 145; Indels 45; Gaps 8;

QY 6 VAVCVVRPLNSREESLG-ETAQVYWKTDNNVIYQVDSKSFNDRVFHGNKTTKNVYEE 64
 Db 10 VRVALRCRPLVSKIEKGCQCLSFVPGEPOVV-VGNDKSFYDFVDPSTQEVEVT 67
 QY 65 IAAPILDSAIQYNGTIFAYGQTASGKTYTMG-----SEDLGVIPRAIHDIFQKIK 117
 Db 68 AVAPLIKGVKGYNAVLAYGQTGSGKTYSMGAYTAEQENBPTGVIPRVIQLLFKVID 127
 QY 118 KPPDRFLLRVSYMEIYNETITDLLCGT-QMKPLIREDVNRNYYVADLTVEEVYTTSEMA 177
 Db 128 KKSDFEFTLVSKYLEIYNEEILDLCPSEKQAQINREDPKGKIVGLTEKTVLVALDT 187
 QY 178 LKWTITKEKSRHYGETKMNORSRSHITFRMILESREKGPSPNCEGSKVSHLNLVDLAG 237
 Db 188 VSCLEQGNNSRTVASTAMNSQSSRSHAITFISIEQRKSD-----KNSSFRSKLHLVDLAG 243
 QY 238 SERAAQTGAAGVRLKEGCNINRSLFILGOVTKLSDGVGGINYRDSKLTILONSLGG 297
 Db 244 SEROKTKAEGDRLEKGININRGLLCLGNVISALGDDKGGFVPRDSKLTILONSLGG 303
 QY 298 NPKTRIITCTPV--SPDETALQFASHTAKYMKNTPYVNEVSTDEALLKRYKREIMDLK 355
 Db 304 NSHTLMTACVSPADSNLEETLNTLYRADRKINKKPIV-----IDPQAEALNHLK 355
 QY 356 KQLEEVSL-----ETRAQMEKQQLAQLEKLLQKQVNEKLENL 396
 Db 356 QVQQLQVLLLAHQGGTLPGSITVRPSENLSIMEKQ--SLVEENKLSRGLSEAAQGT 413
 QY 397 TRML 400
 Db 414 AQML 417

RESULT 10
 KF4A MOUSE
 ID KF4A MOUSE STANDARD; PRT: 1231 AA.
 AC P33174;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Chromosome-associated kinesin KIF4A (Chromokinesin).
 GN KIF4A OR KIF4 OR KNS4.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, AND TISSUE
 RP SPECIFICITY.
 RC STRAIN=ICR; TISSUE=Brain;
 RX MEDLINE=95014709; PubMed=7929562;
 RA Sekine Y., Okada Y., Noda Y., Kondo S., Aizawa H., Takemura R.,
 RA Hirokawa N.:
 RT "A novel microtubule-based motor protein (KIF4) for organelle
 RT transports, whose expression is regulated developmentally";
 RL J. Cell Biol. 127:187-201(1994).

RN SEQUENCE OF 91-240 FROM N.A.
 RP TISSUE=Brain;
 RX MEDLINE=93077686; PubMed=1447303;
 RA Aizawa H., Sekine Y., Takemura R., Zhang Z., Nangaku M.,
 RA Hirokawa N.:
 RT "Kinesin family in murine central nervous system";
 RL J. Cell Biol. 119:1287-1296(1992).
 CC -!- FUNCTION: REQUIRED FOR MITOTIC CHROMOSOMAL POSITIONING AND BIPOLAR
 CC SPINDLE STABILIZATION.
 CC -!- SUBCELLULAR LOCATION: Nuclear. Associated with mitotic
 CC chromosomes.
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN PYRAMIDAL CELLS IN JUVENILE
 CC HIPPOCAMPUS, GRANULAR CELLS IN JUVENILE CEREBELLAR CORTEX AND IN
 CC ADULT SPLEEN.
 CC -!- SIMILARITY: Belongs to the kinesin-like protein family.
 CC Chromokinesin subfamily.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL; D12646; BAA02167.1; --
 CC PIR; A54803; A54803.
 CC HSSP; P17113; 3KAR.
 CC MGD; MGI:108389; Kif4.
 CC InterPro: IPR001752; Kinesin_motor.
 CC Pfam; PF00225; kinesin; 1.
 CC PRINTS; PR00380; KINESINHEAVY.
 CC SMART; SM00129; Kisc; 1.
 CC PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
 CC PROSITE; PS00067; KINESIN MOTOR DOMAIN2; 1.
 CC Motor protein; Microtubule; ATP-binding; DNA-binding;
 CC Nuclear protein; Coiled coil.
 CC DOMAIN 1 350 KINESIN-MOTOR.
 CC FT DOMAIN 351 1000 COILED COIL (BY SIMILARITY).
 CC FT DOMAIN 1001 1231 GLOBULAR.
 CC FT NP BIND 88 95 ATP (POTENTIAL).
 CC FT CONFLICT 112 112 I -> S (IN REF. 2).
 SQ SEQUENCE 1231 AA; 139551 MW; F34F2C2D21158FE4 CRC64;

Query Match 33.8%; Score 693; DB 1; Length 1231;
 Best Local Similarity 39.3%; Pred. No. 1.2e-36;
 Matches 167; Conservative 69; Mismatches 143; Indels 46; Gaps 10;

QY 6 VAVCVVRPLNSREESLG-ETAQVYWKTDNNVIYQVDSKSFNDRVFHGNKTTKNVYEE 64
 Db 10 VRVALRCRPLVSKIEKGCQCLSFVPGEPOVV-VGNDKSFYDFVDPSTQEVEVT 67
 QY 65 IAAPILDSAIQYNGTIFAYGQTASGKTYTMG-----EDH---LGVIPRAIHDIFQKIK 117
 Db 68 AVAPLIKGVKGYNAVLAYGQTGSGKTYSMGAYTAEQENBPTGVIPRVIQLLFKVID 127
 QY 118 KPPDRFLLRVSYMEIYNETITDLLCGT-QMKPLIREDVNRNYYVADLTVEEVYTTSEM 176
 Db 128 KKSDFEFTLVSKYLEIYNEEILDLCPSEKQAQINREDPKGKIVGLTEKTVLVALDT 187
 QY 177 ALKWTITKEKSRHYGETKMNORSRSHITFRMILESREKGPSPNCEGSKVSHLNLVDLAG 236
 Db 188 TVSCLEQGNNSRTVASTAMNSQSSRSHAITFISIEQRKSD-----NDKNSFRSKLHLVDLAG 243
 QY 237 GSERAAQTGAAGVRLKEGCNINRSLFILGOVTKLSDGVGGINYRDSKLTILONSLGG 296
 Db 244 GSEROKTKAEGDRLEKGININRGLLCLGNVISALGDDKGGFVPRDSKLTILONSLGG 303
 QY 297 GNPKTRIITCTPV--SPDETALQFASHTAKYMKNTPYVNEVSTDEALLKRYKREIMDL 354
 Db 304 GNSHTIMTACVSPADSNLEETLNTLYRADRKINKKPIV-----IDPQAEALNHL 355

QY 355 KQLEEVSL-----ETRAQMEKQQLAQLLEKLLQKVQNEKIEN 395
 Db 356 KQVQQLQILLQAHGGLPGDINVERPSNLQSLMEKNQ--SLVEENKLSRGLSEAAQG 413
 QY 396 LTRML 400
 Db 414 TAQWL 418

RESULT 11
 KF17 HUMAN
 ID KF17 HUMAN STANDARD; PRT; 1029 AA.
 AC Q9P2E2; Q95077; Q9N411;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Kinesin-like protein KIF17 (KIF3-related motor protein).
 GN KIF17 OR KIF3X OR KIAA1405.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kimberley A., White S.;
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
 RV [2]
 RP SEQUENCE OF 78-1029 FROM N.A. (ISOFORM 1).
 RC TISSUE=Brain;
 RA Ohara O., Nagase T., Kikuno R.;
 RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
 RV [3]
 RP SEQUENCE OF 239-1029 FROM N.A. (ISOFORM 1).
 RC TISSUE=Brain;
 RX MEDLINE=20181126; PubMed=10718198;
 RA Nagase T., Kikuno R., Ishikawa K.-I., Hirotsawa M., Ohara O.;
 RT "Prediction of the coding regions of unidentified human genes. XVI.
 RT The complete sequences of 150 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RL DNA Res. 7:65-73(2000).
 RV [4]
 RP SEQUENCE OF 618-1029 FROM N.A. (ISOFORM 2).
 RC TISSUE=Pancreas;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.D., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Carninci P., Prange C.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Abramson R.D., Mullaly S.J.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
 RA Blakesley R., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Small D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RV [5]
 RP SEQUENCE OF 788-1029 FROM N.A.
 RA Nagata K., Puls A., Futter C., Aspenstrom P., Hall A.;
 RT "The MAP kinase kinase kinases MLK2 and MLK3 are targets for RAC/Cdc42
 RT and components of microtubule motor complexes.";
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: Transports vesicles containing N-methyl-D-aspartate
 CC (NMDA) receptor 2B along microtubules (By similarity).
 CC -1- SUBUNIT: Interacts with LIN-10 PDZ domain (By similarity).

CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q9P2E2-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q9P2E2-2; Sequence=VSP_008218;
 CC Note=Sequence incomplete. No experimental confirmation
 CC available;
 CC -1- SIMILARITY: Belongs to the kinesin-like protein family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>
 CC or send an email to license@ebi.ac.uk).

DR EMBL; AL663074; -; NOT ANNOTATED CDS.
 DR EMBL; AL391357; -; NOT ANNOTATED CDS.
 DR EMBL; AB037826; BA92643.2; ALT_INIT.
 DR EMBL; BC036871; AAH36871.1; -;
 DR EMBL; AF009624; AA01428.1; -;
 DR HSSP; P33176; 1BG2.
 DR Genew; HGNC:19167; KIF17.
 DR MIM; 605037; -;
 DR InterPro; IPR001752; kinesin_motor.
 DR Pfam; PF00225; kinesin; 1.
 DR SMART; SM00129; KISC; 1.
 DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
 DR PROSITE; PS0067; KINESIN_MOTOR_DOMAIN2; 1.
 KW Motor protein; Microtubule; ATP-binding; Coiled coil; Transport;
 KW Protein transport; Alternative splicing.
 FT DOMAIN 1 265 KINESIN-MOTOR.
 FT DOMAIN 346 462 COILED COIL (POTENTIAL).
 FT DOMAIN 739 846 COILED COIL (POTENTIAL).
 FT NP_BIND 91 98 ATP (POTENTIAL).
 FT VARSPPLIC 619 673 VEAKLRUSVTVARTDAPQADVPKVPQVQVPAFTDLLEPSDA
 FT RPEAAADDFPPRP -> INRACFCVTGVPELEVICLSGLQ
 FT CAEWPRGLSHLGAVGGGGLPTSTWVSLHTQ (in
 FT isoform 2).
 FT /FTid=vsp_008218.
 FT SDEDS -> IGRN (IN REF. 5).
 FT CONFLICT 788 792 MISSING (IN REF. 4).
 FT CONFLICT 908 908 D -> Y (IN REF. 5).
 FT CONFLICT 925 925 R -> W (IN REF. 2 AND 3).
 FT CONFLICT 957 957 THNSPGLSCPLNNSAI -> HTSQLATRPQLTQOQLC
 FT CONFLICT 971 989 H (IN REF. 5).
 SQ SEQUENCE 1029 AA; 115113 MW; EC852E8B4AD18656 CRC64;
 Query Match 33.7%; Score 691.5; DB 1; Length 1029;
 Best Local Similarity 41.0%; Pred. No. 1.1e-36;
 Matches 171; Conservative 62; Mismatches 139; Indels 45; Gaps 11;
 QY 5 AVAVCVVRPLNSRELSIGETAQVYKTD-----NNVIYQVDSKSFNDFRVHNE 56
 Db 5 AVKVVRCRPMNQREEL--RCQPVTVDCARACCCQNFCAADEPPKQFTFDGAYVDH 62
 QY 57 TTKNVYEEIAAPLIDSAIQYNGTIFAYGTASGTVTMMGSED--HLGVIPRAIHDF 113
 Db 63 VTEQIYNEIAYPLVEGTEGYNGTIFAYGTGSGKSTMQGLDPPPSQGRGIIPRAFEHF 122
 QY 114 QTKIKPFDREFLRVSYMEIYNETITDLLCGTQMKPLIIREQVNRNVYVADLTFEYVYT 173
 Db 123 ESVQCAENTKFLVRASYLEIYNEDVRLLGADTKQK-LLEKHEPEKGVYKGLSMHTVHS 181
 QY 174 SEMALKWIKGKSRHYGETKKNQSRSHITFIRMILES---EKGPSNCEGSKVSHL 230
 Db 182 VAOCEHIMETGKNRSVGYTLMNKKSSRSHISFTIETMSAVDERG-----KDLRAGLK 236
 QY 231 NLVDLAGSRAAQTGAAGVRLKEGCGNINSLFILGVIKLSQGVGGFVNYRDSKLTRI 290
 Db 237 NLVDLAGSERQSKTGATGERLKEATKINLSLALGNVISALVDGRC-KHVPYRDSKLTRL 295

```

291 LQNSLGGNPKTRIICTIPV--SFDETILQFASAKYMKNTPPYVNEVSTDEALLKRYR 348
296 LQDSLGGNPKTKLMVACLSPADNNYDETILSTLYRANRAKIRNPRINPKD-ALLREYQ 354
349 KEIMDLKKQLEBVSLETTAAQMEKQIAQLI-----FEKLLQKVQNEKIE 394
355 EIEKKLKAIL-----TQQMSPSSLSALLSRQVPPDPVQVEKLLPQFVIOHME 403

RESULT 12
KF4A_XENLA
ID KF4A_XENLA STANDARD; PRT; 1226 AA.
AC Q21784; Q9PSI0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chromosome-associated kinesin KLP1 (Chromokinesin).
GN KLP1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodidae; Xenopus.
NCBI_TaxID=8355;
[1]
RN SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, AND TISSUE
RP SPECIFICITY.
RC TISSUE=Oocyte.
RX MEDLINE=95236444; PubMed=7720067;
RA Vernos I., Raats J., Hirano T., Heasman J., Karsenti E., Wylie C.;
RT "Xklp1, a chromosomal Xenopus kinesin-like protein essential for
RT spindle organization and chromosome positioning.";
RL Cell 81:117-127(1995).
[2]
RN SEQUENCE OF 9-338 FROM N.A.
RP

```

RT "Multiple kinesin-like transcripts in *Xenopus oocytes*.";
 RL Dev. Biol. 157:232-239 (1993).
 CC -!- FUNCTION: REQUIRED FOR MITOTIC CHROMOSOMAL POSITIONING AND BIPOLAR
 CC SPINDLE STABILIZATION.
 CC -!- SUBCELLULAR LOCATION: Nuclear. Associated with mitotic
 CC chromosomes.
 CC -!- TISSUE SPECIFICITY: Expressed in oocytes, eggs, testes and
 CC brain.
 CC -!- SIMILARITY: Belongs to the kinesin-like protein family.
 CC Chromokinesin subfamily.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

 CC EMBL: X82012; CAA57539.1; --
 DR PIR: I51617; I51617.
 DR HSP: P17119; 3KAR.
 DR InterPro: IPR001752; kinesin_motor.
 DR Pfam: PF00225; kinesin; 1.
 DR PRINTS: PR00380; KINESINHEAVY.
 DR SMART: SM00129; KISC; 1.
 DR PROSITE: PS00411; KINESIN_MOTOR_DOMAIN1; 1.
 DR PROSITE: PS50067; KINESIN_MOTOR_DOMAIN2; 1.
 DR KW Motor protein; Microtubule; ATP-binding; DNA-binding;
 KW Nuclear protein; Coiled coil.
 FT DOMAIN 1 350 KINESIN-MOTOR.
 FT DOMAIN 351 1006 COILED COIL (BY SIMILARITY).
 FT DOMAIN 1007 1226 GLOBULAR.
 FT NP BIND 87 94 ATP (POTENTIAL).
 FT CONFLICT 163 163 I -> L (IN REF. 2).
 FT SEQUENCE 1226 AA; 138923 MW; 7F0275FCF3316697 CRC64;
 SC

[illegible]

Db 300 QSLGNGSHLMIACVSPADSNLETLNLRVADRARKKNKPIVNVDPQAAELNHLKQ 358
 QY 350 EIMDLKKQLEE-----VSLTRAQA-----MKDQLAQLLEKDLLOK-----VQNEK 392
 Db 359 QVQQLQVLLQAHGGLTPVINSAPSENLSILMEKNQ--SLMEENKLSRGLSEAAGQT 416
 QY 393 INENLRLMVT 402
 Db 417 AQMLERIIVT 426

RESULT 15
 KL61_DROME STANDARD; PRT; 1066 AA.
 ID KL61_DROME Q9W01B;
 AC P46863; Q8TOA6; Q9W01B;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Bipolar kinesin KRP-130 (Kinesin-like protein Klp61F).
 GN KLP61F OR KLP2 OR CG9191.
 OS Drosophila melanogaster (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephyarodea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN
 RP SEQUENCE FROM N.A.
 RC MEDLINE=Embryo;
 RX MEDLINE=94043448; PubMed=8227131;
 RA Heck M.M.S., Pereira A., Pesavento P.A., Yannoni Y., Spradling A.C.,
 RA Goldstein L.S.B.;
 RA "The kinesin-like protein KLP61F is essential for mitosis in
 RT Drosophila";
 RL J. Cell Biol. 123:665-679 (1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Deicher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Kimmel B.E., Kodira C., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Riden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Shue B.C., Sengen-Kiamos I., Simpson M., Strong R., Sun E.,
 RA Spier E., Spradling A.C., Stapleton M., Turner R., Wang A.H., Wang X.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S.-C., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195 (2000).
 RN [3]
 RP REVISIONS.
 RX MEDLINE=22426069; PubMed=12537572;
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a
 RT systematic review";
 RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22 (2002).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley; TISSUE=Embryo;
 RX MEDLINE=22426086; PubMed=12537569;
 RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
 RA George R.A., Guarin H., Kronmiller B., Pacle J.M., Park S., Wan K.H.,
 RA Rubin G.M., Celniker S.E.;
 RT "A Drosophila full-length cDNA resource";
 RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8 (2002).
 RN [5]
 RP SEQUENCE OF 228-357 FROM N.A.
 RC STRAIN=DP CN BW;
 RX MEDLINE=92020874; PubMed=1924306;
 RA Stewart R.J., Pesavento P.A., Woerpel D.N., Goldstein L.S.B.;
 RT "Identification and partial characterization of six members of the
 RL kinesin superfamily in Drosophila";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:8470-8474 (1991).
 RN [6]
 RP PARTIAL SEQUENCE, AND IDENTIFICATION AS KRP-130.
 RX MEDLINE=97078747; PubMed=8918872;
 RA Kashina A.S., Scholey J.M., Leszyk J.D., Saxton W.M.;
 RT "An essential bipolar mitotic motor";
 RL Nature 384:225-225 (1996).
 CC -!- FUNCTION: IMPORTANT ROLE IN MITOTIC DIVIDING CELLS. MICROTUBULE
 CC MOTOR REQUIRED FOR SPINDLE BODY SEPARATION. SLOW PLUS-END DIRECTED
 CC MICROTUBULE MOTOR CAPABLE OF CROSS-LINKING AND SLIDING APART
 CC ANTIPARALLEL MICROTUBULES. THEREBY PUSHING APART THE ASSOCIATED
 CC SPINDLE FOLDS DURING SPINDLE ASSEMBLY AND FUNCTION.
 CC -!- SUBUNIT: Homotetramer.
 CC -!- DEVELOPMENTAL STAGE: SPECIFICALLY EXPRESSED IN PROLIFERATING
 CC TISSUES DURING EMBRYONIC AND LARVAL DEVELOPMENT.
 CC -!- PTM: PHOSPHORYLATION DURING MITOSIS AT THR-933 CONTROLS THE
 CC ASSOCIATION OF KLP61F WITH THE SPINDLE APPARATUS (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. BIMC
 CC SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U01842; AAA03718.1; -
 CC EMBL; AB003471; AAF47458.2; -
 CC EMBL; AY069442; AAL39587.1; -
 CC EMBL; M74428; AAA28655.1; -
 CC PIR; A48669; A48669.
 CC HSP; P17119; 3KAR.
 CC FlyBase; FBgn004378; Klp61F.
 CC GO; GO:0005737; Cytoplasm; IDA.
 CC GO; GO:0005874; Cytoskeleton; IDA.
 CC GO; GO:0003774; F-motor activity; IDA.
 CC GO; GO:0007100; P-centrosome separation; IGI.
 CC InterPro; IPR001752; kinesin_motor.

Search completed: July 29, 2004, 09:36:15
Job time : 4.88047 secs

GenCore version 5.1.6

Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 09:27:31 ; Search time 17.1603 Seconds
(without alignments)
7428.155 Million cell updates/sec

Title: US-10-045-631B-88_COPY_2_405

Perfect score: 2053

Sequence: 1 AEGAVAVCVRVPLNSREE.....LOKVQNKIENLRMLVTSS 404

Scoring table:

BLOSOM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25.*

1: sp_archaea.*

2: sp_bacteria.*

3: sp_fungi.*

4: sp_human.*

5: sp_invertebrate.*

6: sp_mammal.*

7: sp_mhc.*

8: sp_organelle.*

9: sp_phage.*

10: sp_plant.*

11: sp_rodent.*

12: sp_virus.*

13: sp_vertebrate.*

14: sp_unclassified.*

15: sp_ivirus.*

16: sp_bacteriaph.*

17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1799	87.6	549	11	Q7TPX4	Q7tpx4 mus musculus
2	1468.5	71.5	2954	13	O42263	O42263 xenopus lae
3	872	42.5	807	10	Q94HV9	Q94hv9 arabidopsis
4	872	42.5	823	10	Q9S7P3	Q9s7p3 arabidopsis
5	856	41.7	459	10	Q9SS30	Q9s830 arabidopsis
6	836.5	40.7	888	10	Q9LO62	Q9lq62 arabidopsis
7	784	38.2	1885	5	Q869B8	Q869b8 dictyosteli
8	758.5	36.9	1055	10	Q8RW44	Q8rw44 arabidopsis
9	758.5	36.9	1055	10	Q8W5R5	Q8w5r5 arabidopsis
10	754	36.7	1459	3	P87198	P87198 ustilago ma
11	754	36.7	2013	5	Q9VKI0	Q9vki0 drosophila
12	754	36.7	2244	5	Q9NCG0	Q9ncg0 drosophila
13	748.5	36.5	1033	10	Q9LHL9	Q9lhl9 arabidopsis
14	748.5	36.5	1058	10	Q9SUJ0	Q9sju0 arabidopsis
15	727.5	35.4	956	10	Q9C7B9	Q9c7b9 arabidopsis
16	726	35.4	443	11	Q8CGJ1	Q8cgj1 mus musculus

17	724	35.3	997	10	Q9FG03	Q9fg03 arabidopsis
18	723	35.2	701	11	Q7TSZ7	Q7tsz7 mus musculus
19	722.5	35.2	959	10	Q8S950	Q8s950 nicotiana t
20	721.5	35.1	890	10	Q8W5R6	Q8w5r6 arabidopsis
21	721	35.1	699	5	Q8MPT8	Q8mpt8 caenorhabdi
22	720	35.1	408	11	Q8OUK1	Q8ouk1 mus musculu
23	716.5	34.9	408	4	Q8IWH8	Q8iwh8 homo sapien
24	715.5	34.9	954	10	Q9AMW8	Q9amw8 oryza sativ
25	714.5	34.8	699	13	Q98T11	Q98t11 xenopus lae
26	713	34.7	747	11	Q8BNH4	Q8brh4 mus musculu
27	713	34.7	757	11	Q8OU27	Q8ou27 mus musculu
28	707	34.4	160	11	O3S059	O3s059 mus musculu
29	706.5	34.4	1193	10	Q7X7H8	Q7x7h8 oryza sativ
30	704	34.3	677	5	Q9VRK9	Q9vrk9 drosophila
31	704	34.3	782	5	Q19633	Q19633 caenorhabdi
32	704	34.3	782	5	Q9GV92	Q9gv92 caenorhabdi
33	703	34.2	735	5	Q9U0D5	Q9u0d5 tetrahymena
34	702	34.2	974	10	Q8S905	Q8s905 arabidopsis
35	701	34.1	836	5	Q7YUC7	Q7yuc7 tetrahymena
36	695	33.9	1037	10	Q9FH58	Q9fh58 arabidopsis
37	694.5	33.8	671	4	Q86XX7	Q86xx7 homo sapien
38	694.5	33.8	1127	4	Q86TN3	Q86tn3 homo sapien
39	694	33.8	1121	10	Q9SVI8	Q9svi8 arabidopsis
40	693	33.8	819	11	Q7TQG6	Q7tqg6 mus musculu
41	693	33.8	1231	11	Q80YP3	Q80yp3 mus musculu
42	692	33.7	909	10	Q9XI03	Q9xi03 arabidopsis
43	687.5	33.5	727	3	Q86Z96	Q86z96 gibberella
44	686.5	33.4	581	10	Q9SUJ7	Q9sju7 arabidopsis
45	686	33.4	671	5	Q8MPT7	Q8mpt7 caenorhabdi

ALIGNMENTS

RESULT 1

Q7TPX4 PRELIMINARY; PRT; 549 AA.

ID Q7TPX4
AC Q7TPX4;
DT 01-OCT-2003 (TRENBLrel. 25, Created)
DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)
DE 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]

SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Egg;

RX MEDLINE=22388257; PubMed=12477932;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Buetow K.H., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,

RA Brownstein M.J., Udén T.B., Toshyuki S., Carninci P., Prange C.,

RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,

RA Krzywinski M.I., Skalska J., Smailus D.E., Schnerch A., Schein J.E.,

RA Jones S.J., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length human

RT and mouse cDNA sequences.";

RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RP [2]

RC SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Egg;

RA Strausberg R.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
KW EMBL; BC052843; AAHS2843.1; -.
KW Hypothetical protein.
FT NON_TER 549 549
SQ SEQUENCE 549 AA; 62899 MW; B707C97DC9FB38D3 CRC64;

Query Match 87.6%; Score 1799; DB 11; Length 549;
Best Local Similarity 87.1%; Pred. No. 2.9e-116; Indels 0; Gaps 0;
Matches 352; Conservative 17; Mismatches 35;

QY 1 AEEGAVAVCVVRPLNSREESLGETAQVYKTKDNNVIYQVDSKSFNFVRVHGNETTNN 60
DB 2 AEEASVAVCVVRPLNSREELGEATHIYKTKDKNALYQSDGKSFQDFRVFDSNETTKN 61

QY 61 VVEEIAAPIDSIAIOGNGTIFAYGQTASGKTYTMGSEDHLGVIPRAIHDIPOKIKKFP 120
DB 62 VVEEIAVPIISSAIOGNGTIFAYGQTASGKTYTMGSEDCLGVIPRAIHDIPOKIKKFP 121

QY 121 DREFLARVSMETINETITDLGCTOKMKPLIREDVNRVYVADLTEEVVYVYSEMAK 180
DB 122 EREFLARVSMETINETITDLGCTOKMKPLIREDVNRVYVADLTEEVVYVYSEMAK 181

QY 181 ITKGEKSRHYGKTMNORSRSHITPRMILESREKGEPSNCEGSKVSHNLVLDLAGSER 240
DB 182 LATGEKRNHYGKTMNORSRSHITPRMILESREKGEPSNCEGSKVSHNLVLDLAGSER 241

QY 241 AAGTGAAGVRLKGCNCNINRSILFQGVIKKSDGVGFFINRDSKLTILQNSLGNPK 300
DB 242 AAGTGAAGVRLKGCNCNINRSILFQGVIKKSDGVGFFINRDSKLTILQNSLGNPK 301

QY 301 TRICTITPVSFDETLTALOPASTAKYKNTPPYNEVSTDEALKRYRKEIMDLKQLEE 360
DB 302 TRICTITPVSFDETLTALOPASTAKYKNTPPYNEVSTDEALKRYRKEIMDLKQLEE 361

QY 361 VSLTEFRAQAMEKDLQALKEKLLQKQVNEKIENLRLMLVTSS 404
DB 362 VNTKTRAQAMEKDLQALKEKLLQKQVNEKIENLRLMLVTSS 405

RESULT 2
ID O42263 PRELIMINARY; PRT; 2954 AA.
AC O42263
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Kinesin-related protein.
GN XCENP-E.

OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
[1]

RP SEQUENCE FROM N.A.
RX MEDLINE=98028574; PubMed=9363944;
RA Wood K.W., Sakowicz R., Goldstein L.S., Cleveland D.W.;
RT "CENP-E is a plus end-directed kinetochore motor required for
metaphase chromosome alignment.";
RL Cell 91:357-366 (1997).

DR EMBL; AF027728; AAC60300.1; -.
DR PIR; T14156; T14156.
DR HSP; F33176; 1B2.
DR GO; GO:0005871; C:kinesin complex; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR GO; GO:0007017; F:microtubule-based process; IEA.
DR InterPro; IPR001752; kinesin_motor.
DR Pfam; PF00225; kinesin; 1.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.

DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
SQ SEQUENCE 549 AA; 62899 MW; B707C97DC9FB38D3 CRC64;

DR PROSITE; PS00067; KINESIN_MOTOR_DOMAIN2; 1.
SQ SEQUENCE 2954 AA; 339964 MW; 439804ED0E592679 CRC64;

Query Match 71.5%; Score 1468.5; DB 13; Length 2954;
Best Local Similarity 70.9%; Pred. No. 2.3e-92; Indels 7; Gaps 3;
Matches 290; Conservative 52; Mismatches 60;

QY 1 AEEGAVAVCVVRPLNSREESLGETAQVYKTKDNNVIYQVDSKSFNFVRVHGNETTNN 60
DB 2 SEGDAVKCVVRPLTQREQ--GDQANLQWKAGNNTISQVDTGKSFNFVRVHGNETTNN 59

QY 61 VVEEIAAPIDSIAIOGNGTIFAYGQTASGKTYTMGSEDHLGVIPRAIHDIPOKIKKFP 120
DB 60 IQEETAVPIIRALQGYNGTIFAYGQTSSGKTYTMGTSNLSGLIIPQAIQEVFKIIQIEIP 119

QY 121 DREFLARVSMETINETITDLGCTOKMKPLIREDVNRVYVADLTEEVVYVYSEMAK 180
DB 120 NREFLARVSMETINETITDLGCTOKMKPLIREDVNRVYVADLTEEVVYVYSEMAK 179

QY 181 ITKGEKSRHYGKTMNORSRSHITPRMILESREKGEPSNCEGSKVSHNLVLDLAG 237
DB 180 IKGEKRNHYGKTMNORSRSHITPRMILESREKGEPSNCEGSKVSHNLVLDLAG 239

QY 238 SERAAGTGAAGVRLKGCNCNINRSILFQGVIKKSDGVGFFINRDSKLTILQNSLGG 297
DB 240 SERAAGTGAAGVRLKGCNCNINRSILFQGVIKKSDGVGFFINRDSKLTILQNSLGG 299

QY 298 NPRTIICITTPVSFDETLTALOPASTAKYKNTPPYNEVSTDEALKRYRKEIMDLKQ 357
DB 300 NAKTVIICITTPVSFDETLTALOPASTAKYKNTPPYNEVSTDEALKRYRKEIMDLKQ 359

QY 358 LE-EVSLTEFRAQAMEKDLQALKEKLLQKQVNEKIENLRLMLVTSS 404
DB 360 LENLESSETKAQAMEKDLQALKEKLLQKQVNEKIENLRLMLVTSS 408

RESULT 3
Q94HV9 PRELIMINARY; PRT; 807 AA.
ID Q94HV9
AC Q94HV9;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Kinesin motor protein (kin2), putative.
GN T4M14.11.

OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
[1]

RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Town C.D., Haas B.J., Wu D., Maiti R., Hannick L.I., Chan A.P.,
RA Tallon L.J., Rooney T., Utterback T.R., Vanaken S.E., Feidblyum T.V.,
RA White O., Fraser C.M.;
RT "Arabidopsis thaliana chromosome 1 BAC T4M14 genomic sequence.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AC027036; AAK62792.1; -.
DR GO; GO:0005871; C:kinesin complex; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR GO; GO:0007017; F:microtubule-based process; IEA.
DR InterPro; IPR001752; kinesin_motor.
DR Pfam; PF00225; kinesin; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR PROSITE; PS00067; KINESIN_MOTOR_DOMAIN2; 1.
SQ SEQUENCE 807 AA; 91260 MW; Fd7CDAD68EA30C28 CRC64;

Query Match 42.5%; Score 872; DB 10; Length 807;
Best Local Similarity 47.6%; Pred. No. 7.6e-52;


```

RX MEDLINE=98359834; PubMed=9693369;
RA de Hostos E.L., McCaffrey G., Sugang R., Pierce D.W., Vale R.D.;
RT "A developmentally regulated kinesin-related motor proteins from
RT Dictyostelium discoideum.";
RL Mol. Biol. Cell 9:2093-2106(1998).
RN [2]
RN SEQUENCE FROM N.A.
RP Suyama E., Sutoh K.;
RA "Kinesin-related proteins from Dictyostelium.";
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB102780; BAC56912.1; -
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005875; C:microtubule associated complex; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR GO; GO:0001752; kinesin_motor.
DR InterPro; IPR001752; kinesin_motor.
DR Pfam; PF00225; kinesin; 1.
DR Pfam; PF02370; M; 6.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR PROSITE; PS00667; KINESIN_MOTOR_DOMAIN2; 1.
SQ SEQUENCE 1885 AA; 218216 MW; 15686ABD3B007EC7 CRC64;

Query Match 38.2%; Score 784; DB 5; Length 1885;
Best Local Similarity 41.4%; Pred. No. 3.1e-45;
Matches 183; Conservative 81; Mismatches 124; Indels 54; Gaps 10;

Qy 2 EGVAVCVVRPLNSREESLGETAQYWKTDNNVI-YQVDGSKSFNDRVFNHNETKN 60
Db 19 ELNKKIKVAIRVPLNSRE-LGIDQKIPWSIKDTISLSQNPINFYDVFGLDSTID 76
Qy 61 VYIEIAPIIDSALQGNITFAYGQTASGTYTMGSEDLGVIPRAIHDFOKIK-KF 119
Db 77 VYNAIAKSIIVNSLNGINGTIFAYGQTSSGKTFSMRGTESIPGIKLSIKDIFKSDSI 136
Qy 120 PDREFLAVSMYELNETITDLCGT-QKMKPLIREDVNRNVVADLTERRVYVTSMAL 178
Db 137 LEKDYLLKVSYLEIYNEIKDNLNPTISNKKKIKHEDIYKGVVAVANLKEEIVISPOIF 196
Qy 179 KWIPTGKSRHYGTQKNQSRSHITPRMILESRKGEPSNCGSVKVSHLNLDLAGS 238
Db 197 ALNFGGEERHIGSTMWDSRSSRSHITFRMQIQSTCKQ-----NGTIQMSTLTLDLAGS 251
Qy 239 ERAAQTGAAGVRLKGCNINRSFILGOVIKLSGQGVGFINTRDSKLTILQNSLGGN 298
Db 252 ERVSTGAEGVRLKGEHINKSLMTLSKVISKLSSEKTTQOHVPYRDSKLTILQPSLGGN 311
Qy 299 PKTRICTITPVSF--DETLTALQFASAKYMKNTPPYVNEVSTDEALLKRYRKEIMDLKK 356
Db 312 SKTALCTITPATTHQESISTLOFAKRAKRVKYNKINQVADANTMLKKYSEILELQON 371
Qy 357 QL-----EEVSL-----ETRAQAMEKDL--AQ 377
Db 372 QLVKSEERINSLRTISQEISSNNFKLGMRKFNDAIIGSLINENKKKKRRNTLDPSY 431
Qy 378 LLEEKDLLQKV---QNEKIENL 396
Db 432 LLKDKIINKKIRKGENQKIKKI 453

RESULT 8
QBRWW4 PRELIMINARY; PRT; 1055 AA.
AC QBRWW4;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative kinesin.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```

```

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OX eurosids II; Brassicales; Brassicaceae; Arabidopsais.
OX NCBI_TaxID=3702;
RN [1]
RN SEQUENCE FROM N.A.
RP Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M.,
RA Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,
RA Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G.S., Bowser L.,
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Shinozaki K.,
RA Davis R.W., Ecker J.R., Theologis A.;
RT "Arabidopsis Full Length cDNA Clones.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY091060; AM13891.1; -
DR GO; GO:0005871; C:kinesin complex; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR GO; GO:0007017; P:microtubule-based process; IEA.
DR InterPro; IPR001752; kinesin_motor.
DR InterPro; IPR001841; Znf_ring.
DR Pfam; PF00225; kinesin; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR SMART; SM00184; RING; 1.
DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR PROSITE; PS00667; KINESIN_MOTOR_DOMAIN2; 1.
DR PROSITE; PS50089; ZF_RING_2; 1.
SQ SEQUENCE 1055 AA; 116462 MW; BBF3709ACB40215B CRC64;

Query Match 36.9%; Score 758.5; DB 10; Length 1055;
Best Local Similarity 42.0%; Pred. No. 8e-44;
Matches 176; Conservative 84; Mismatches 136; Indels 23; Gaps 8;

Qy 1 AEGAVAVCVVRPLNSREESLGETAQYWKTDNNVIYQVDGSKSFNDRVFNHNETKN 60
Db 94 SERDSISVTVRPLNSREYQRGDEVAWYPDGTDLVRHEYNPLTAYAFDKVFGPQATID 153
Qy 61 VYIEIAPIIDSALQGNITFAYGQTASGTYTMGSEDLGVIPRAIHDFOKIK-KF 120
Db 154 VYDVAAPVVKAAMEGVNITFAYGVTSSGKTHTMHGDQESPGHIPLAKDVFSIIQDTP 213
Qy 121 DREPLLRVSMYELNETITDLCGTQKMKPLIREDVNRNVVADLTERRVYVTSMALKW 180
Db 214 GRELLAVSYLEIYNEIVINDLLDPTG--QNLRVRED--SQGTVEGKEEVLSLPGHALSF 270
Qy 181 ITGKESRHYGTQKNQSRSHITPRMILESRKGEPSNCGSVKVSHLNLDLAGSER 240
Db 271 IAAGEEHRHVGSNFNLLSSRSHTIFTLMVESSATGDEYD---GVTFSQLNLDLAGSE- 326
Qy 241 AAQTGAAGVRLKGCNINRSFILGOVIKLSGQGVGFINTRDSKLTILQNSLGGNPK 300
Db 327 SSKTETTLTKRRKESYINKSLTLTGTVIGLSEK-ATHIPYRDSKLTILQNSLGGH 385
Qy 301 TRICTITPV--SPDETTLTALQFASAKYMKNTPPYVNEVSTDEALLKRYRKEIMDLKKQL 358
Db 386 VSLICTITPASSSEETHNTLKFSRAKSIHFIYASRNQIIDEKSLIKKYQREISTLKLEL 445
Qy 359 EEV-----SLETRAQAMEKDL--LAQLLEKDLLQKQNEKIENTRMLVTS 404
Db 446 DQLRRGMLGVVSHBELMSLKQLEEGQVQKQSRLEEBEAKAALMSRIQKLTILVST 504

RESULT 9
QBRWW5 PRELIMINARY; PRT; 1055 AA.
ID QBRWW5;
AC QBRWW5;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Kinesin-related protein (Putative kinesin protein).
GN MKRP2 OR AT4G39050.

```

OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP MEDLINE=21563048; PubMed=11706156;
 RX Itoh R., Fujiwara M., Yoshida S.;
 RA "Kinesin-Related Proteins with a Mitochondrial Targeting Signal";
 RT Plant Physiol. 127:724-726(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Hsuan V.W., Lee J.M.,
 RA Quach H.L., Tang C.C., Toriumi M., Wallender E.K., Wong C., Wu H.C.,
 RA Yu G., Yuan S., Chen H., Cheuk R., Jones T., Kim C.J., Nguyen M.,
 RA Palm C.J., Shinn P., Southwick A., Tripp M.G., Wu T., Davis R.W.,
 RA Ecker J.R., Theologis A.;
 RT "Arabidopsis Open Reading Frame (ORF) Clones";
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB062739; BAB71852.1; -.
 DR EMBL; AX150516; AAN13032.1; -.
 DR GO; GO:0005871; C:kinesin complex; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0003774; F:motor activity; IEA.
 DR GO; GO:0007017; P:microtubule-based process; IEA.
 DR InterPro; IPR001752; P:microtubule-based process; IEA.
 DR InterPro; IPR001841; Znf ring.
 DR Pfam; PF00225; kinesin; 1.
 DR PRINTS; PR00380; KINESINHEAVY.
 DR SMART; SM00129; KISC; 1.
 DR SMART; SM00184; KINC; 1.
 DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
 DR PROSITE; PS0067; KINESIN MOTOR DOMAIN2; 1.
 DR PROSITE; PS0089; ZF RING 2; 1.
 SQ SEQUENCE 1055 AA; 116463 MW; B69EC383FF14AB7B CRC64;
 Query Match 36.9%; Score 758.5; DB 10; Length 1055;
 Best Local Similarity 42.0%; Pred. No. 8e-44;
 Matches 176; Conservative 84; Mismatches 136; Indels 23; Gaps 8;
 QY 1 AEEGAVAVRVPLNSREESLGETAQVYKTDN--NVYQVD----- 41
 DB 94 SERDSISVTVRPLSDREVQRGDEVAWTPDGLVRRHEYNPLTAYAFKDPGQATTID 153
 QY 61 VYEEIAAIPDLSAIOGYNGTIFAYGOTASGKTYMMGSEDHLGVIPIRAIHDFQKIKFP 120
 DB 154 VYDVAARPVVKAAMEGVNGTVFAYGVTSSGKTHTMHGDOESPGLIPLAIKDVFSIIQDTP 213
 QY 121 DREFLLRVSYMEYNETITDLCGTOKMKPLIREDVNRNVYVADLTVEVYVTSSEALKW 180
 DB 214 GREFLRVSYLEYNEVINDLDPTG--QNLVRVED--SQGTVEGKEEVYVLSPGHALSF 270
 QY 181 ITKGEKSRHVGETKMNORSRSHITFRMILESREKGEPCNCEGVKVVSHNLVDLAGSER 240
 DB 271 IAGEEHRHVGNFNFLSRSHITFTLVVSSATGDEYD---GVTFSQLNLDLAGSE 326
 QY 241 AAQTGAAGVRLKEGCNINRSLFILGQVKKLSQGVGFINYRDSKILTRILQNSLGNPK 300
 DB 327 SSKTETTGLRRKEGYSINKSLTLTGTVIGKLSGK-ATHIPYRDSKILTRILQSSLSGHG 385
 QY 301 TRIICITITPV--SFDITLALQPASTAKYKNTPTPYNEVSTDEALKRYRKEIMDLKKQL 358
 DB 386 VSLICTITPASSESTHNTLTPASRAKSGIEIYASRNQIIDEKSLIKKYQREISTJLKL 445
 QY 359 EEV-----SLETRAQAMEKQD--LAQLEEKDLQKQVNEKIENLTRLMTSS 404
 DB 446 DQLRRGMVGVSHLEMLSLKQLEEGQVKKQSELEEEBAKALMSRIKILKILIVST 504
 RESULT 10
 P87198
 ID P87198 PRELIMINARY; PRT; 1459 AA.

AC P87198;
 DT 01-JUL-1997 (TrEMBLrel. 04, Created)
 DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Kinesin motor protein.
 GN KIN1.
 OS Ustilago maydis (Smut fungus).
 OC Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;
 OC Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.
 OX NCBI_TaxID=5270;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97361828; PubMed=9218789;
 RA Lehmler C., Steinberg G., Snetelaar K.M., Schliwa M., Kahmann R.,
 RA Bolker M.;
 RT "Identification of a motor protein required for filamentous growth in
 Ustilago maydis";
 RL EMBL J. 16:3464-3473(1997).
 DR EMBL; U28844; AAB63336.1; -.
 DR PIR; T30196; T30196.
 DR HSSP; P17119; 3KAR.
 DR GO; GO:0005871; C:kinesin complex; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0003774; F:motor activity; IEA.
 DR GO; GO:0007017; P:microtubule-based process; IEA.
 DR InterPro; IPR001752; kinesin; 1.
 DR Pfam; PF00225; kinesin; 1.
 DR PRINTS; PR00380; KINESINHEAVY.
 DR SMART; SM00129; KISC; 1.
 DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
 DR PROSITE; PS0067; KINESIN MOTOR DOMAIN2; 1.
 SQ SEQUENCE 1459 AA; 156932 MW; 281B74B0100F5CDO CRC64;
 Query Match 36.7%; Score 754; DB 3; Length 1459;
 Best Local Similarity 38.2%; Pred. No. 2.6e-43;
 Matches 190; Conservative 70; Mismatches 127; Indels 110; Gaps 14;
 QY 6 VAVCVVRPLNSREESLGETAQVYKTDN--NVYQVD----- 41
 DB 237 VVVCVRMRP--SRASSDSEASV-WNCDSEKNRIFPTEHHPALAKRTTSSERAGASIA 293
 QY 42 -----GSKSFNDRVPHGNETTKNVYEEAAPLIDSAIOGYNGTIFAYGQ 86
 DB 294 AAPSSHDLDHEDPTSSYTHYFQDKLITGAQTTDDMYHSIAPVVRVAEINGTIVFAYGQ 353
 QY 87 TAGSKTYTMGSEDHLGVIPIRAIHDFQKIKFPDEFLLRVSYMEYNETITDLCGTQ 146
 DB 354 TSGSKTHTWGSDAEPGVIPIRAVEQIFQMIKDEPDEFLLRVSYLEYNETLKDLLA--- 410
 QY 147 KMKPL-----IREDVNRNVYVADLTVEV 171
 DB 411 PLPPLTGTSGSLQTTDRPASPIKGGSSHAAGSQSCTLRILIEDQKSSRVIITGLREEIV 470
 QY 172 YTSSEALKWITKGEKSRHVGETKMNORSRSHITFRMILESREKGEPCNCEGVKVVSHLN 231
 DB 471 TANTVLCUIQRGQDRRHVGATDWNERSRSHCVFOLTIESRPPAPSASKE--VRISQLN 528
 QY 232 LVDLAGSERAAGTGAAGVRLKEGCNINRSLFILGQVKKLSQGVGF--FINVRDSKLTR 289
 DB 529 LIDLAGSERAA---SQAEERKEGAFINKSLTLTGTVIGKLTPEVNGDAHIPIYRDSKLTR 585
 QY 290 ILQNSLGNPKTRIICITITPV--SFDITLALQPASTAKYKNTPTPYNEVSTDEALKRY 347
 DB 586 ILQTSLSGNARIAVICTLSPDTHANETLSLTKFGKRCVKLVVTTAKKGTAMDMDKALQKY 645
 QY 348 RKEIMDLKKQLEE-----VSLERAQAMEK--DOLAOLLE-----EKDLLQKVQ 389
 DB 646 RRELDALRAKLKANGSPNPNEVMTIVSAESKESQOKLDQLNQKEAAQREVEDMQKCR 705
 QY 390 NE---KIENLTRLMTSS 403
 DB 706 SHLKAQIEHLTRLILTS 722

Query Match 36.7%; Score 754; DB 5; Length 2013;

Db	8	STQVCIKVRPCPLTSL-----WQVKEGRSIIHLADSHABPVYFDYVDFGASQNEVF	60
QY	63	BEIAPIDSAIOGYNCTIFAYGQTASGKTYTMMGSEDHLGVIPRAIHDFQIKKFPDR	122
Db	61	DRMARHIVACMGQNGTIFAYGQTSSGKTYTMMGDEQNGPVWVLAKEIFQOISSETER	120
QY	123	BEILRVSYMEIYNETITDILGCGTKMKPILIRDNVNRNVYVADLTEEVVYTSMAKWI	181
Db	121	DFLVRGVYIIEYNEKIYDIL--NKKQDLKIHESGNGVNVN--CEEIITSEVDLLRL	176
QY	182	TKGEKSRHGETKONQSSRSHTIFRMILESRKGPSPNCEGSKVSHLNVLDAGSERA	241
Db	177	CLNGKERTVGETNWNERSRSHAIFKIIIESR-KSDHSD-DDAVIOSVNLVLDAGSERA	234
QY	242	AQTGAACVRLKEGNCINRSFILGOVTKLSDGVGGFINVRDSKLTIRLQNSLGGNPKT	301
Db	235	DOTGARGARLKEGHINKSLFLSNVTKSUSENADNFTNRSKLTIRLQNSLGGNAFT	294
QY	302	RIITCTTPVSFDETLTALQFASAKYMKNTFYVNEVSTDEALLKRYRKEIMDLKKOLEEV	361
Db	295	SIITCTIPSMESQSLSFATRAKKIRIKPQVNMVSDATMMKRLEREIKVLKDLAE--	353
QY	362	SLETRAQAMKQDLAQLLEKDLQVNEKIEHLEML	400
Db	354	-----EER---KNENQKQVEHLERQI	371
Db	397	TRMLVTSS	404
Db	478	TKLIVST	485
Db	70	VTVVFRFLSPREIRQGE--EVAWYADGETIVRNEHNPTIAYADRVFGPTTTTNRVVD	127
QY	64	EIAAPIIDSAIOGYNCTIFAYGQTASGKTYTMMGSEDHLGVIPRAIHDFQIKKFPDR	123
Db	128	IAAHVWVNGAMEGINTIFAYGVTSSGKTHTHMGDQSPGIIPLAVKDAFSIIQETPNE	187
QY	124	FLLRVSMEIYNETITDILGCGTKMKP-----LIREDVNRNVYVADLTEEVVYTSMAK	179
Db	188	FLLRVSMEIYNEVNDLL-----NPAGNHLRIED-KQGTVEGIEKEEVLSAPAHLS	240
QY	180	WITKGE---KSRHYGETKONQSSRSHTIFRMILESRKGPSPNCEGSKVSHLNVLDLA	236
Db	241	LIAAGEVMTQEHVGSNENLLSSRGHTIFTIETSSPLGDKGKE-AVHLSQNLVLDLA	299
QY	237	GSRAAQTAGVRLKEGNCINRSFILGOVTKLSDGVGGFINVRDSKLTIRLQNSL	296
Db	300	GSE-SGKVTSGVRRKEGYSINKSLTGLTVISKLTIDVR-ASHVPFRDSKLTIRLQNSLS	357
QY	297	GNPKTRITCTIPV--SFDETLTALQFASAKYMKNTFYVNEVSTDEALLKRYRKEIMDL	354
Db	358	GHDRVSLICVTTPASSSEETHNTLFAHRAKHIEIQAEQNKTIIDEKSLIKKYQREIRQL	417
QY	355	KKOLEEVSLETRAQAMEKD-----QLAQLLEKDLQVNEKIEHLEML	396
Db	418	KEELEQLKQEIIVPVPQLKDIGADDIVLLKQLEKQVGLQSRLEEEEBEAKALLSRQL	477
QY	397	TRMLVTSS	404
Db	478	TKLIVST	485
Db	995JU0	PRELIMINARY; PRT; 1058 AA.	
AC	Q9S5JU0	Q9S5JU0; Q94BQ1;	
DT	01-MAY-2000	(TrEMBLrel. 13, Created)	
DT	01-JUN-2002	(TrEMBLrel. 21, Last sequence update)	
DT	01-OCT-2003	(TrEMBLrel. 25, Last annotation update)	
DE		Purative kinesin heavy chain.	
GN	AT2G21380.		
OS	Arabidopsis thaliana (Mouse-ear cress).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;		
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsids.		
OX	NCBI_TaxID=3702;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Cv. Columbia;		
RA	Lin X., Kaul S., Shea T.P., Fujii C.Y., Shen M., VanAken S.E.,		
RA	Barnstead M.E., Mason T.M., Bowman C.L., Ronning C.M., Benito M.-I.,		
RA	Carrera A.J., Creasy T.H., Buell C.R., Town C.D., Nierman W.C.,		
RA	Fraser C.M., Venter J.C.;		
RL	Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RA	Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,		
RA	Dale J.M., Gibson H.A., Goldsmith A.D., Jiang P.X., Lee J.M.,		
RA	Quach H.L., Tang C., Toriumi M., Yu G., Bowser L., Carninci P.,		
RA	Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T., Kamiya A.,		
RA	Karlin-Neumann G., Kawai J., Kim C., Koeseema E., Lam B., Lin J.,		
RA	Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,		
RA	Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Tracy S.E.,		
RA	Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;		
RT	"Full Length cDNA of gene F3K23.14/At2G21380 (GI:4567271).";		
RL	Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RA	Yamada K., Chan M.M., Chang C.H., Dale J.M., Hsuan V.W., Lee J.M.,		
RA	Quach H.L., Tang C., Toriumi M., Wallender E.K., Wong C., Wu H.C.,		
RA	Yu G., Yuan S., Carninci P., Chen H., Cheuk R., Hayashizaki Y.,		
RA	Ishida J., Jones T., Kamiya A., Kawai J., Kim C.J., Narusaka M.,		
RA	Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shinn P.,		

RESULT 13

Db	Q9LHL9	PRELIMINARY; PRT; 1033 AA.	
AC	Q9LHL9;		
DT	01-OCT-2000	(TrEMBLrel. 15, Created)	
DT	01-OCT-2000	(TrEMBLrel. 15, Last sequence update)	
DT	01-OCT-2003	(TrEMBLrel. 25, Last annotation update)	
DE		Kinesin (Centromere protein) like heavy chain-like protein.	
OS	Arabidopsis thaliana (Mouse-ear cress).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;		
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsids.		
OX	NCBI_TaxID=3702;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Columbia;		
RA	Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;		
RL	Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Columbia;		
RX	MEDLINE=20363099; PubMed=10907853;		
RA	Nakamura Y.;		
RT	"Structural analysis of Arabidopsis thaliana chromosome 3. II.		
RT	Structure features of the regions of 4,251,695 bp covered by ninety P1,		
RT	TAC and BAC clones.";		
RL	DNA Res. 7:217-221(2000).		
DR	EMBL; AP002040; BAB03114.1; -.		
DR	HSPB; P33176; 1BG2.		
DR	GO; GO:0005871; C:kinesin complex; IEA.		
DR	GO; GO:0005524; P:ATP binding; IEA.		
DR	GO; GO:0003774; F:motor activity; IEA.		
DR	GO; GO:0007017; P:microtubule-based process; IEA.		
DR	InterPro; IPR001752; P:kinase motor.		
DR	PFAM; PF00225; kinesin; 1.		
DR	PRINTS; PR00380; KINESINHEAV.		
DR	SMART; SM00129; KTSC; 1.		
DR	PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.		
DR	PROSITE; PS50067; KINESIN MOTOR DOMAIN2; 1.		
SQ	SEQUENCE 1033 AA; 116718 MW; 741F5977FD6048FB CRC64;		
Query Match	36.5%; Score 748.5; DB 10; Length 1033;		
Best Local Similarity	41.8%; Pred. No. 3.9e-43;		
Matches 179; Conservative	77; Mismatches 131; Indels 41; Gaps 11;		
QY	6	VAVCVRVRLNSREESLGETAQVYVKTNNVIVQDGSK--SFNDFRVHGNETKNVYE	63

[illegible]

Fri Aug 6 10:49:11 2004

QY 358 LEEVLETRAQAMEYD-----OLAQLLEKDLLOKVONEKINLITRM 399
||:| |
Db 409 LEQLKOEIVVPQLXDIGADDIVLLKQKLEDCQVKLQSRLEEEEAALLSRIQRLTKL 468
||:| |
QY 400 LVTSS 404
||:| |
Db 469 ILVST 473

Search completed: July 29, 2004, 09:39:41
Job time : 18.1603 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 09:26:01 ; Search time 27:9291 Seconds
(without alignments)
4694.096 Million cell updates/sec

Title: US-10-045-631b-88_COPY_2_465

Perfect score: 2360

Sequence: 1 AREGAVAVCVVRPLNSREE.....ITTKTKLSINLLRDEISV 464

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match: 0%

Maximum Match: 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04.*

1: Geneseqp1980s.*

2: Geneseqp1990s.*

3: Geneseqp2000s.*

4: Geneseqp2001s.*

5: Geneseqp2002s.*

6: Geneseqp2003as.*

7: Geneseqp2003bs.*

8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2360	100.0	2633	4	ABG06505	Abg06505 Novel hum
2	2360	100.0	2663	4	AAM39097	Aam39097 Human pol
3	2251.5	95.4	2688	4	AAM40883	Aam40883 Human pol
4	1549	65.6	2954	2	AA01632	Aay01632 Amino aci
5	820	34.7	366	4	ABU53125	Abu53125 Intracell
6	781	33.1	348	4	ABU53208	Abu53208 Human cel
7	778.5	33.0	2013	4	ABBE2322	Abb62322 Drosophil
8	716.5	30.4	677	4	ABBE5183	Abb65183 Drosophil
9	704	29.8	1029	5	AAE17786	Aael17786 Human kin
10	704	29.8	1029	5	ADCI10190	Adci10190 Human NOV
11	697.5	29.6	1232	7	ADDA49938	Adda49938 Human lun
12	697.5	29.6	1232	7	ADDI18924	Addi18924 Human dis
13	695	29.4	473	5	ABG70992	Abg70992 Human tar
14	695	29.4	522	5	ABG70991	Abg70991 Human HSK
15	695	29.4	1038	5	AAM48337	Aam48337 Human kin
16	694.5	29.4	1048	4	ABBS9245	Abbs9245 Drosophil
17	694.5	29.4	1066	4	AAAG67418	Aag67418 Amino aci
18	694.5	29.4	1066	6	ABG72693	Abg72693 Fruitfly
19	694	29.4	1034	3	AAAG31112	Aag31112 Arabidops
20	694	29.4	1063	3	AAAG31111	Aag31111 Arabidops
21	694	29.4	1121	3	AAAG31110	Aag31110 Arabidops
22	688.5	29.2	1234	5	ABG70993	Abg70993 Human HSK
23	686.5	29.1	1232	7	ADDA49937	Adda49937 Human kin
24	686.5	29.1	1232	7	ADDA49932	Adda49932 Human lun
25	680.5	28.8	1232	5	ABG70990	Abg70990 Human HSK

RESULT 1
ABG06505
ID ABG06505 standard; protein; 2633 AA.

XX AC ABG06505;

DT DT 13-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #6496.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX FN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US008631.

XX PR 31-MAR-2000; 2000US-00540217.

XX PR 23-AUG-2000; 2000US-00649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX N-PSDB; AAS70692.

XX PT New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensic, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity.

XX PS Claim 20; SEQ ID NO 36864; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging

ALIGNMENTS

Adb67093 Kinesin h
Aau19569 Human dia
Abp51294 Human MDD
Ade55349 Rat Prote
Aaw72746 Drosophil
Adb63485 Drosophil
Adb67088 Kinesin h
Aaw72745 Drosophil
Aaw72744 Drosophil
Aau74840 Human HSK
Aae14400 Human kin
Aau79590 Human kin
Abr48222 Human bla
Adb80468 Ovarian c
Adc35116 Human bre
Abp68930 Human pol
Abg60124 Human DIT
Abb2962 Drosophil
Adb67091 Kinesin h
Aag31117 Arabidops

Fri Aug 6 10:49:11 2004

CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 2633 AA;

Query Match 100.0%; Score 2360; DB 4; Length 2633;
Best Local Similarity 100.0%; Pred. No. 1.9e-182; Indels 0; Gaps 0;
Matches 464; Conservative 0; Mismatches 0

QY 1 AEEGAVAVCVVRPLNSREESLGETAQVYWKTDNNVIYQVDSKSFNDRVPHGNETTKN 60
DB 2 AEEGAVAVCVVRPLNSREESLGETAQVYWKTDNNVIYQVDSKSFNDRVPHGNETTKN 61
QY 61 VVEEIAAPIIDSALIQYNGTIFAYGQTASGKTYTMGSEDLHGVIPRAIHDFQIKKFP 120
DB 62 VVEEIAAPIIDSALIQYNGTIFAYGQTASGKTYTMGSEDLHGVIPRAIHDFQIKKFP 121
QY 121 DREFLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNYYVADLTVEEVYTSMAK 180
DB 122 DREFLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNYYVADLTVEEVYTSMAK 181
QY 181 ITKGEKSRHYGETKMNORSRSHITFRMILESREKGEPSNCEGSKVYSHLNLVDLAGSR 240
DB 182 ITKGEKSRHYGETKMNORSRSHITFRMILESREKGEPSNCEGSKVYSHLNLVDLAGSR 241
QY 241 AAQTGAAGVRLKEGCNINRSFLILGVIKKLSGQGVGFNDRVPHGNETTKN 300
DB 242 AAQTGAAGVRLKEGCNINRSFLILGVIKKLSGQGVGFNDRVPHGNETTKN 301
QY 301 TRIICTITPVSFDETLTALQFASTAKYMNTPYNEVSTDEALLKRYRKEIMDLKKOLEE 360
DB 302 TRIICTITPVSFDETLTALQFASTAKYMNTPYNEVSTDEALLKRYRKEIMDLKKOLEE 361
QY 361 VSLTRAQAMEKDQALQLEEKDLQVNEKINLTMLVTSSSLTLQQLKAKRKRV 420
DB 362 VSLTRAQAMEKDQALQLEEKDLQVNEKINLTMLVTSSSLTLQQLKAKRKRV 421
QY 421 TWCLGKINKMNSYADQFNIPNTITTKTKLSINLREIDESV 464
DB 422 TWCLGKINKMNSYADQFNIPNTITTKTKLSINLREIDESV 465

RESULT 2
AAM39097
ID AAM39097 standard; protein; 2663 AA.
XX AAM39097;
AC AAM39097;
XX 22-OCT-2001 (first entry)
XX Human polypeptide SEQ ID NO 2242.
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX leukaemia.
XX Homo sapiens.
XX WO200153312-A1.
XX 26-JUL-2001.
PD

XX 26-DEC-2000; 2000WO-US034263.
XX 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
PR 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
PR 29-NOV-2000; 2000US-00727344.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang J, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao Qa;
PI Zhou P, Goodrich R, Drmanac RT;
XX WPI; 2001-442253/47.
DR N-PSDB; AAI58253.
XX Novel nucleic acids and polypeptides, useful for treating disorders such
XX as central nervous system injuries.
XX Example 4; SEQ ID NO 2242; 10078pp; English.
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and the
XX encoded polypeptides (AAM38642-AAM42213) with nootropic,
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localised neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: Immune system suppression,
XX Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX assays for receptor activity, arthritis and inflammation, leukaemias and
XX C.N.S disorders. Note: The sequence data for this patent did not form
XX part of the printed specification

XX Sequence 2663 AA;
SQ Query Match 100.0%; Score 2360; DB 4; Length 2663;
Best Local Similarity 100.0%; Pred. No. 1.9e-182; Indels 0; Gaps 0;
Matches 464; Conservative 0; Mismatches 0

QY 1 AEEGAVAVCVVRPLNSREESLGETAQVYWKTDNNVIYQVDSKSFNDRVPHGNETTKN 60
DB 2 AEEGAVAVCVVRPLNSREESLGETAQVYWKTDNNVIYQVDSKSFNDRVPHGNETTKN 61
QY 61 VVEEIAAPIIDSALIQYNGTIFAYGQTASGKTYTMGSEDLHGVIPRAIHDFQIKKFP 120
DB 62 VVEEIAAPIIDSALIQYNGTIFAYGQTASGKTYTMGSEDLHGVIPRAIHDFQIKKFP 121
QY 121 DREFLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNYYVADLTVEEVYTSMAK 180
DB 122 DREFLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNYYVADLTVEEVYTSMAK 181
QY 181 ITKGEKSRHYGETKMNORSRSHITFRMILESREKGEPSNCEGSKVYSHLNLVDLAGSR 240
DB 182 ITKGEKSRHYGETKMNORSRSHITFRMILESREKGEPSNCEGSKVYSHLNLVDLAGSR 241
QY 241 AAQTGAAGVRLKEGCNINRSFLILGVIKKLSGQGVGFNDRVPHGNETTKN 300
DB 242 AAQTGAAGVRLKEGCNINRSFLILGVIKKLSGQGVGFNDRVPHGNETTKN 301
QY 301 TRIICTITPVSFDETLTALQFASTAKYMNTPYNEVSTDEALLKRYRKEIMDLKKOLEE 360
DB 302 TRIICTITPVSFDETLTALQFASTAKYMNTPYNEVSTDEALLKRYRKEIMDLKKOLEE 361

QY 361 VSLTEAQAAMEKDQQLAQLLBEKDLLOKQVQNEKIENLTRLMLVTSSSLTLOQELKAKKRRV 420
 DB 362 VSLTEAQAAMEKDQQLAQLLBEKDLLOKQVQNEKIENLTRLMLVTSSSLTLOQELKAKKRRV 421
 QY 421 TWCLGKINKMKNKNYADQFNIPNTITTKTKLSINLLREIDSV 464
 DB 422 TWCLGKINKMKNKNYADQFNIPNTITTKTKLSINLLREIDSV 465

RESULT 3

ID AAM40883

XX AAM40883 standard; protein; 2688 AA.

AC AAM40883;

XX 22-OCT-2001 (first entry)

DE Human polypeptide SEQ ID NO 5814.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW leukinaemia.

XX Homo sapiens.

OS WO200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US034263.

XX 23-DEC-1999; 99US-00471275.

XX 21-JAN-2000; 2000US-00488725.

XX 25-APR-2000; 2000US-00523317.

XX 20-JUN-2000; 2000US-00598042.

XX 19-JUL-2000; 2000US-00620312.

XX 03-AUG-2000; 2000US-00653450.

XX 14-SEP-2000; 2000US-00662191.

XX 19-OCT-2000; 2000US-00693036.

XX 29-NOV-2000; 2000US-00727344.

XX (HYSE-) HYSE INC.

XX Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren P, Wang D;

XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;

XX Zhou P, Goodrich R, Drmanac RT;

XX WPI; 2001-442253/47.

XX N-PSDB; AAI60039.

XX Novel nucleic acids and polypeptides, useful for treating disorders such
 XX as central nervous system injuries.

XX Example 2; SEQ ID NO 5814; 10078pp; English.

XX The invention relates to human nucleic acids (AAI57798-AAI61369) and the

XX encoded polypeptides (AAM38642-AAM42213) with nootropic,

XX immunosuppressant and cytostatic activity. The polynucleotides are useful

XX in gene therapy. A composition containing a polypeptide or polynucleotide

XX of the invention may be used to treat diseases of the peripheral nervous

XX system, such as peripheral nervous injuries, peripheral neuropathy and

XX localised neuropathies and central nervous system diseases, such as

XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic

XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the

XX utilisation of the activities such as: Immune system suppression,

XX Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic

XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,

XX assays for receptor activity, arthritis and inflammation, leukaemias and

XX C.N.S disorders. Note: The sequence data for this patent did not form

XX part of the printed specification

XX

SQ Sequence 2688 AA;

Query Match

Best local similarity 95.4%; Score 2251.5; DB 4; Length 2688;

Matches 451; Conservative 2; Mismatches 11; Indels 3; Gaps 3;

QY 1 AEGAVAVCVVRPLNSREESLGETAQVYWKTDNNVIYQVDGSKSNFDFVFGNETTKN 60

DB 23 AEGAVAVCVVRPLNSREESLGETAQVYWKTDNNVIYQVDGSKSNFDFVFGNETTKN 82

QY 61 VVEETIAAPIIDSAIQYNGTIFAYGQTASGKTVMGSEDHGLGVIPIRA-IHDIF-QKIKK 118

DB 83 VVEETIAAPIIDSAIQYNGTIFAYGQTASGKTVMGSEDHGLGVIPIQGFHGFHSQKIXE 142

QY 119 -FPDREFLLRVSYMEIYNETITDLCGTQKMKPLIIRVDNRNRYVADLTTEEVTYSEMA 177

DB 143 VFLDREFLLRVSYMEIYNETITDLCGTQKMKPLIIRVDNRNRYVADLTTEEVTYSEMA 202

QY 178 LKWIITGKESRHYGETKMKQRSRSHITFRMILESEKGEPCNCEGSKVSHLNVLVDLAG 237

DB 203 LKWIITGKESRHYGETKMKQRSRSHITFRMILESEKGEPCNCEGSKVSHLNVLVDLAG 262

QY 238 SERAAQTGAAGVRLKEGNCNINRSFLTGQVVKLSGQVGGFVNYRDSKLTRLQNSLGG 297

DB 263 SERAAQTGAAGVRLKEGNCNINRSFLTGQVVKLSGQVGGFVNYRDSKLTRLQNSLGG 322

QY 298 NPKTRIICTITPVSFDETLTALQFASTAKYKNTPVNEVSTDEALLKRYRKEIMDLKKQ 357

DB 323 NPKTRIICTITPVSFDETLTALQFASTAKYKNTPVNEVSTDEALLKRYRKEIMDLKKQ 382

QY 358 LEEVSLETRAQAMEKDQQLAQLLBEKDLLOKQVQNEKIENLTRLMLVTSSSLTLOQELKAKRK 417

DB 383 LEEVSLETRAQAMEKDQQLAQLLBEKDLLOKQVQNEKIENLTRLMLVTSSSLTLOQELKAKRK 442

QY 418 RRVTWCLGKINKMKNKNYADQFNIPNTITTKTKLSINLLREIDSV 464

DB 443 RRVTWCLGKINKMKNKNYADQFNIPNTITTKTKLSINLLREIDSV 489

RESULT 4

AAAY01632

ID AAAY01632 standard; protein; 2954 AA.

XX AAAY01632;

XX 22-JUN-1999 (first entry)

XX Amino acid sequence of centromere-associated protein-E (CENP-E).

CENP-E; centromere-associated protein-E; ATPase activity;

plus end-directed microtubule motor activity; chromosome congression;

microtubule binding activity; chromosome movement; mitosis;

cell proliferation; tumor; metastasis; vascular malfunction;

inflammatory disease; immune disease; angiogenesis; hypertension;

restenosis; fungal infection; selective herbicide; fungicide;

insecticide; plant growth regulator; activator; cancer cell marker.

XX Xenopus sp.

OS WO9913061-A1.

XX 18-MAR-1999.

XX 10-SEP-1998; 98WO-US019231.

XX 11-SEP-1997; 97US-0058645P.

XX (REGC) UNIV CALIFORNIA.

XX Wood KW, Sakowicz R, Goldstein LSB, Cleveland DW;

XX WPI; 1999-229233/19.

XX PN WO200112659-A2.
XX PD 22-FEB-2001.
XX PF 18-AUG-2000; 2000WO-IB001496.
XX PR 18-AUG-1999; 99US-0149499P.
XX PR 28-SEP-1999; 99US-0156503P.
XX PA (GEHU-) GERMAN HUMAN GENOME PROJECT.
XX PI Wiemann S;
XX XX WPI; 2001-327840/34.
XX DR Nucleic acids having the sequences of clones isolated from libraries of
XX PT different human tissues, useful in recombinant DNA methodologies.
XX XX Example III; Page 745; 1095pp; English.
XX PS This invention describes novel polynucleotides and polypeptides isolated
XX CC from human cDNA libraries which can be used for gene therapy or in
XX CC vaccines. The polynucleotides of the invention and antibodies encoded by
XX CC them may be used in the prevention, diagnosis and treatment of diseases
XX CC associated with inappropriate polypeptide expression. The products of the
XX CC invention may also be used to identify modulators of expression and
XX CC activity and to down regulate expression and activity. The antibodies of
XX CC the invention may also be used as diagnostic agents for detecting the
XX CC presence of polypeptides in samples. This sequence represents a homologue
XX CC of a polypeptide described in the disclosure of the invention
XX SQ Sequence 366 AA;
Query Match 34.7%; Score 820; DB 4; Length 366;
Best Local Similarity 50.5%; Pred. No. 3.4e-58;
Matches 187; Conservative 42; Mismatches 115; Indels 26; Gaps 7;
Qy 11 RVRPLNREBSLGTAAQYVW-----KTDNNVIYQVDGSKSFNDRVPHGNETTKNVYEE 64
Db 1 RCRLNREINDGSCVVPWPPTGYKTVHNGHGDSPHKSFTFDHVFVWNCIOEDVYDT 60
Qy 65 IAAPIIDSAIGYNGTIFAYGQTASGKTYTWG-----SEHLGVIPRAIHDFQIKKFP 120
Db 61 VAHIVDDCFYNGCTIFAYGQTGSKTYTWGPGGHEPDHMGIIIPRCHDIFDRIDKFQ 120
Qy 121 DRE---FLLRVSYMEIYNETITDLLC-GTQCKPLIREDVNRNYYVADLTVEVYVYSEM 176
Db 121 EKDHDFVHVKCSYMEIYNEEIVDLLCPNPQHMKPLNIHEHPNMGYVQGCTEFFHVCSYED 180
Qy 177 ALKWTKEKSRHYGETKQNRSSRSHITFRMILESREKBPSPNCEGSKVKVSHLNLVDLA 236
Db 181 ACHWITQGNKRHVAATNMNDHSRSHITFIHVEQRHK-----QDHEVCHSKMNLVDLA 236
Qy 237 GSERAAQTGAAGVRLKEGNCINRSIFILGOVVKKLSDGQV-----GGFINYRDSKLPRI 290
Db 237 GSERVNRFGAGQRLKEGNCINQSITLGNVINALADGQTKMYGCHGHIPIYRDSKLAWL 296
Qy 291 LQNSLGGNPKRIITITPV--SPDETALQFASAKYMKNTPYVNEVSTDEALLKVR 348
Db 297 LQDSLGGNCKTMIACIWPADNVEETLSTLRADRAKNIKNKPOINDPCAMALWRRYH 356
Qy 349 KEIMDLKKQL 358
Db 357 EQIQDMKHQL 366
RESULT 6
ABUS3208
ID ABUS3208 standard; protein; 348 AA.
XX AC ABUS3208;
XX XX

DR N-PSDB; AAX26819.
XX PT Centromere-associated protein-E and related nucleic acid.
XX PS Claim 5; Page 66-67; 77pp; English.
XX XX The present sequence represents CENP-E (centromere-associated protein-E)
XX CC of Xenopus. The protein has at least one of plus end-directed microtubule
XX CC motor activity, Arpase (adenosine triphosphatase) activity and
XX CC microtubule binding activity. CENP-E is the motor that powers chromosome
XX CC movement toward microtubule plus ends and is essential for congression of
XX CC chromosomes during mitosis. Modulators of CENP-E can thus control cell
XX CC proliferation. Agents that modulate CENP-E activity are lead therapeutic,
XX CC bioagricultural and diagnostic agents, e.g. for treatment of unwanted
XX CC cell proliferation (typical of many examples are tumors and metastases;
XX CC vascular malfunction; inflammatory and immune diseases; angiogenesis;
XX CC hypertension; restenosis; and fungal infections), also as plant-
XX CC protection agents (selective herbicides, fungicides and insecticides) and
XX CC plant growth regulators or activators for improving yields. CENP-E is
XX CC also a diagnostic marker for dividing cells, including cancer cells
XX SQ Sequence 2954 AA;
Query Match 65.6%; Score 1549; DB 2; Length 2954;
Best Local Similarity 66.0%; Pred. No. 2.7e-116;
Matches 312; Conservative 63; Mismatches 86; Indels 12; Gaps 5;
Qy 1 AEGAVAVCVVRPLNREBSLGTAAQYVWKTDDNNVIYQVDGSKSFNDRVPHGNETTKN 60
Db 2 SEGDAVVCVRVRLIQREQ--GDQANLQWKAAGNTISQVDGTSFNDRVFNHSESTQ 59
Qy 61 VYEEIAPIIDSAIGYNGTIFAYGQTASGKTYTWGSEDLGVIPRAIHDFQIKKFP 120
Db 60 IYQEIAPVIRLSALQNGTIFAYGQTSSGKTYTWGTPNSLGIIPQAIQEVKLIQBP 119
Qy 121 DRELLRVSYMEIYNETITDLLCGTQCKMKPLIREDVNRNYYVADLTVEVYVYSEMALKW 180
Db 120 NRELLRVSYMEIYNETVKDLLCDRRKKPLEIREDFNRNYYVADLTVEVYVYVYSEM 179
Qy 181 ITKGEKSRHYGETKQNRSSRSHITFRMILESREKGEPS---NCEGSKVKVSHLNLVDLAG 237
Db 180 IKGGEKSRHYGETKQNRSSRSHITFRMIVSRORNDPTNENCDGAVVSHLNLVDLAG 239
Qy 238 SERAAQTGAAGVRLKEGNCINRSIFILGOVVKKLSDGQGVGFYINRDSKLPRIQNSLGG 297
Db 240 SERASQTGAAGVRLKEGNCINRSIFILGOVVKKLSDGQAGGFYINRDSKLPRIQNSLGG 299
Qy 298 NPKTRIICTITPVSPDETTLALQFASAKYMKNTPYVNEVSTDEALLKRYRKEIMDLKKQ 357
Db 300 NAKTVIICITITPVSPDETTLQFASAKYMKNTPYVNEVSTDEALLKRYRKEIMDLKKQ 359
Qy 358 LE--EVSLETRAQAMEKQQLAEKDLQKQVNEKTEMLVTSSTLQOELKAK 415
Db 360 LENLESSETQAQAKAEHTQLAEIKQLHREDEDRILWHLNIVVASSQES-QDQVRK 418
Qy 416 RKRRTVWGLKGNKMSYADQFN---IPTNITTKTKLSINLLREIDESV 464
Db 419 RKRRTVWPGKQNSLHSGVSDFDMLSLPGNFSKAKAFSDMPSPFEDDSV 471
RESULT 5
ABUS3125
ID ABUS3125 standard; protein; 366 AA.
XX AC ABUS3125;
XX XX 15-APR-2003 (first entry)
XX DE Intracellular trafficking-associated DKFphtes3_26g22 homologue #2.
XX XX Human; gene therapy; vaccine; disease treatment; detection.
XX XX Homo sapiens.
OS

DT 14-APR-2003 (first entry)
 DE Human cell cycle-associated DKFzptes3_35b4 homologue #7.
 DE Human; gene therapy; vaccine; disease treatment; detection.
 XX Homo sapiens.

XX WO200112659-A2.
 XX 22-FEB-2001.

XX 18-AUG-2000; 2000WO-IB001496.
 XX 18-AUG-1999; 99US-0149499P.
 XX 28-SEP-1999; 99US-0156503P.

XX (GEHU-) GERMAN HUMAN GENOME PROJECT.
 XX Wiemann S;
 XX WPI; 2001-327840/34.

XX Nucleic acids having the sequences of clones isolated from libraries of
 XX different human tissues, useful in recombinant DNA methodologies.

XX Example III; Page 829; 1095pp; English.

XX This invention describes novel polynucleotides and polypeptides isolated
 XX from human cDNA libraries which can be used for gene therapy or in
 XX vaccines. The polynucleotides of the invention and antibodies encoded by
 XX them may be used in the prevention, diagnosis and treatment of diseases
 XX associated with inappropriate polypeptide expression. The products of the
 XX invention may also be used to identify modulators of expression and
 XX activity and to down regulate expression and activity. The antibodies of
 XX the invention may also be used as diagnostic agents for detecting the
 XX presence of polypeptides in samples. This sequence represents a homologue
 XX of a polypeptide described in the disclosure of the invention

XX Sequence 348 AA;

Query Match 33.1%; Score 781; DB 4; Length 348;
 Best Local Similarity 51.1%; Pred. No. 4.8e-55;
 Matches 178; Conservative 38; Mismatches 106; Indels 26; Gaps 7;
 QY 11 RVRPLNSREESLGETAQVW-----KTDNNVIYQVGSKFNEDEVFHGNETTNNVYEE 64
 DB 1 RCRPLNEREINDGSCVQVWPWTGYKTVHNGHGDSPHKSFTDHFVWNNCTQSDVDYT 60
 QY 65 TAAPIDSALQYNGTIFAYGQTASGKTYTMG-----SEDLHGVIPRAIHDFQIKKFP 120
 DB 61 VAHPVDDCFHGYNCTIFAYGQTGSKTYTMGPGEHPDMGIIIPRCHDIDFDRDKFO 120
 QY 121 DRE--FLLRVSYMEIYNETITDLIC-GTQKKPLIIRDVNRNVYVADLTBEVYVTSEM 176
 DB 121 EKDHDFWFKVCSMEIYNEEYDLLCPNPQHKPLNIHEHPNNGPYVQCTEFHVCSEYD 180
 QY 177 ALKWITKESKSHYGETKMNQSRSSHTIFRMILESRKEGSPNCEGSKYVSHLNLVLA 236
 DB 181 ACHWIAQGNKRNHVATNNHSSSHSTIFTIHVEQRHK-----QCDEHVCHSKMNLVLA 236
 QY 237 GSERAAQTGAAGVRLKEGNCINRSFILQGVIKKLSDGQV-----GGFINVRDLSKLTRI 290
 DB 237 GSERNVRTGAEQRLKEGNCINQSLTTLGNVINALADQTKMYGHHGHIIPRDSKLTWL 296
 QY 291 LQNSLGGNPKRIITITPV--SFDETLTALQFASHTAKYMKNTPYVNE 336
 DB 297 LQDSLGGNCKTMIACIWPADWNYEETLSTLRYADRANKIKNKPQINE 344

RESULT 7

ABB62322

ID ABB62322 standard; protein; 2013 AA.

XX ABB62322;
 XX 26-MAR-2002 (first entry)
 XX Drosophila melanogaster polypeptide SEQ ID NO 13758.
 DE Drosophila; developmental biology; cell signalling; insecticide;
 XX pharmaceutical.

XX Drosophila melanogaster.
 XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;
 XX WPI; 2001-656860/75.

XX N-PSDB; ABL06425.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
 XX genes from Drosophila and for elucidating cell signalling and cell-cell
 XX interactions.
 XX Disclosure; SEQ ID NO 13758; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
 XX capable of detecting 1000 or more genes from Drosophila. The invention is
 XX useful in developmental biology and in elucidating cell signalling and
 XX cell-cell interactions in higher eukaryotes for the development of
 XX insecticides, therapeutics and pharmaceutical drugs. The invention
 XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 XX sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
 XX ABB72072). The sequence data for this patent did not form part of the
 XX printed specification, but was obtained in electronic format directly
 XX from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 2013 AA;

Query Match 33.0%; Score 778.5; DB 4; Length 2013;
 Best Local Similarity 44.3%; Pred. No. 9.3e-54;
 Matches 190; Conservative 66; Mismatches 124; Indels 49; Gaps 12;

QY 5 AVAVCVVRPLNSREESLGETAQVWTKT-DNNVIYQVDP-SKSFNDFRVHGNETTNNVY 62
 DB 8 STQVCIKVRPCBPGLTSL-----WQVKRRSIHLADSHAPYVDFVDFDGSANQEVF 60

QY 63 BEIARPIIDSALQYNGTIFAYGQTASGKTYTMGSEDLHGVIPRAIHDFQIKKFPDR 122
 DB 61 DRMAKHIVHACQMGNGTIFAYGQTSSGKTYTMGDEQNPQVWVLAKEIFQOISSETER 120

QY 123 EFLLRVSYMEIYNETITDLICGTQKKPLIIRDVNRNVYVADLTBEVYVTSEM-ALKWI 181
 DB 121 DFLRVGVIEIYNEKIYDLL--NKNQDLKHESNGIWNVN--CEECIIITSEVDLLRL 176

QY 182 TGKSKSRHYGETKMNQSRSSHTIFRMILESRKEGSPNCEGSKYVSHLNLVLAGSERA 241
 DB 177 CLGNKERTVGETNMNERSRSHAFKIIIESR-KSDHSD-DDAVIQSVNLNVLVLAGSERA 234

QY 242 AQTGAAGVRLKEGNCINRSFILQGVIKKLSDGQVGGFINTRDSKLTFRILQNSLGGNPKT 301
 DB 235 DOTGARGARLKEGGHINKSLFLSNVKSLENADNRFNTYRDSKLTFRILQASLGGNAFT 294

QY 302 RIICITIPVSFDETLTALQFASHTAKYMKNTPYVNEVSTDEALLKRYRKEINDLKKQLEEV 361

Db 295 SIITIKPSIMESQSTSPATRAKKIRIKPQNVMSVDAIMMKRLEREIKVUKDLAE- 353

QY 362 SLETRAQAMEKDQALLREKLLQKQVQNEKIENLTRM-----LVTSSSLTQOELKA 414

Db 354 -----EER-----KQENQKVHLEIRQIKHDMKIKICGHSLS-----DKG 388

QY 415 KKKRRVTWC 423

Db 389 QKKRRVTWC 397

RESULT 8

ABB65183

ID ABB65183 standard; protein; 677 AA.

XX

AC ABB65183;

XX

XX

DT 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster polypeptide SEQ ID NO 22341.

XX

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.

XX

OS Drosophila melanogaster.

XX

XX W0200171042-A2.

XX

PD 27-SEP-2001.

XX

XX 23-MAR-2001; 2001WO-US009231.

XX

XX 23-MAR-2000; 2000US-0191637P.

PR 11-JUL-2000; 2000US-00614150.

XX

XX (PBKE) PE CORP NY.

FA

PI Venter JC, Adams M, Li PWD, Myers EW;

PI

XX WPI; 2001-656860/75.

DR

DR N-PSDB; ABL09286.

XX

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signaling and cell-cell

PT interactions.

XX

XX Disclosure; SEQ ID NO 22341; 2lpp + Sequence Listing; English.

PS

XX

CC The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-

CC ABB72072). The sequence data for this patent did not form part of the

CC printed specification, but was obtained in electronic format directly

CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX

SQ Sequence 677 AA;

Query Match 30.4%; Score 716.5; DB 4; Length 677;

Best Local Similarity 40.8%; Pred. No. 2.2e-49;

Matches 189; Conservative 73; Mismatches 166; Indels 35; Gaps 13;

QY 6 VAVCVVRPLNRSRESLGETAQVYWKTDNNVIQV-----DGSKSFNFDRVHGNETT 58

Db 21 VRVVRTRPMKNDKNSAGLSAISVDKINRAITVMKPNATANEPPKTYTFDNDVFGGSNQ 80

QY 59 KNVYELAAPIIDLSALQGVNGTIFAYGQTASGKTYTMMGSEDH---LGVIPRAIHDFQK 115

Db 81 MDLYVDTRAPIVDKVLGNGTILAYGQTGKTYTMSGNPDSPQTKIIPNAFAHIFGH 140

QY 116 IKKFPD-REFLRLSVMEIYNETITDLCQTKQMKPLIREDVNRNVYVADLTVEVYVS 174

Db 141 IAKAKENQFLVSVMEIYNEVROLL-GKDVGKSLVKERPDIGVFVKDLSGYVYVNA 199

QY 175 EMALKWITTKGKSRHYGETQNRSSRSHTIPFMILESRKKGPSNCEGSKVSHLNVLD 234

Db 200 DDLENIMRLGNKRAVGATQKQNSSRSHAIFISITVERSELGEGD--VQHVRMGKLQQLVD 257

QY 235 LAGSERAAGTGAAGVRLKEGCNINRSILFGLQVVIKKLSDGVGFGFNYRDSKLTRILQNS 294

Db 258 LAGSERQSKTQASGQRLKEATKINLSLVGNVISALVDGK-STHPIYRNSKILTRLLQDS 316

QY 295 LGGNPKTRIITTPV--SFDETTLTALQFASTAKYMKNTPYVNEVSTDEALLKRYRKEIM 352

Db 317 LGGSKTVMCATISPADSNYMETISILRYASRAKNIQNRMHINEEPKD-ALLRHFOSEIA 375

QY 353 DLKKQLEE-VSLETRAQAMEKDQALLREKLLQKQVQNEKIENLTRLMTVSSSLTQO- 410

Db 376 RLKQLEEGDLSLEEPSSE-----EBEDTADDELEAPLE-----TELESSTIQAV 421

QY 411 ELKAKRRVTWC-LGKINKMKNVADQFNIPNTITTKTKL 452

Db 422 EKPKKKREKTDAAKEELAKRKNQEHQKEIEHAKTEQETLRNKL 464

RESULT 9

AAE17786

ID AAE17786 standard; protein; 1029 AA.

XX

AC AAE17786;

XX

DT 07-MAY-2002 (first entry)

XX

DE Human kinesin superfamily motor protein, HsKif17.

XX

XX Human; kinesin superfamily motor protein; HsKif17; autoimmune disease;

KW cellular proliferation; therapy; cancer; hyperplasia; graft rejection;

KW cardiac hypertrophy; immune disorder; inflammation; immunosuppressive;

KW angioplasty; arthritis; restenosis; inflammatory bowel disease;

KW cytostatic; vasotropic; antiarthritic; antiinflammatory.

XX

OS Homo sapiens.

XX

XX W0200198314-A2.

PN

XX 27-DEC-2001.

PD

XX 20-JUN-2001; 2001WO-US019811.

PF

XX 20-JUN-2000; 2000US-00597602.

PR

XX (CYTO-) CYTOKINETICS INC.

PA

XX Beraud C, Freedman R;

PI

DR WPI; 2002-147789/19.

DR

DR N-PSDB; AAD28568.

XX

XX Novel human kinesin superfamily motor protein, HsKif17, useful for

PT identifying modulators that may be used to treat cell proliferative

PT disorders, e.g., cancer, hyperplasias, restenosis, autoimmune disease,

PT arthritis, graft rejection.

XX

PS Claim 11; Fig 2; 64pp; English.

XX

CC The present invention relates to an isolated novel human kinesin

CC superfamily motor protein, HsKif17. HsKif17 is useful for screening

CC modulators of HsKif17, comprising contacting biologically active HsKif17

CC with a candidate agent in a test and control concentration and assaying

CC for the level of HsKif17 activity where a change in activity between the

CC test and control concentration indicates a modulator, and where screening

CC occurs in a multiwell plate as a part of high-throughput screen. The

CC modulators identified are useful in treatment of cellular proliferation

Matches 185; Conservative 75; Mismatches 165; Indels 68; Gaps 15;

QY 5 AVAVCVVRPLNSREESLGETAQVYWKTD-----NNVIYQVDGSKSFNDRVFHGN 56

Db 5 AVKVVRCRPMNQREEL--RCQVTVVDCARACQCIQNPAADEPKQFTFGAYVDH 62

QY 57 TTKNVYEIAAALIDSAIQNGYNTIFAYGQTASGKTYTMMGSD---HLGVIPRAIHDF 113

Db 63 VTREIYNEIAYPLVEGTEGYNITIFAYGQTSGSKFTMQGLPDPSPQSQRGIIIPRAEHF 122

QY 114 QKIKKPPDFEFLRVSYMEIYNETITDLCGTOKMKPLIREDVNRVNYVADLTEEVTY 173

Db 123 ESVQCAENTKFLVRSYLEIYNEIDVRLDGLGADTKQK-LELKEHPEKGVYVKGUSMTVHS 181

QY 174 SEMALKWITKGEKSRHYGETKMQORSRSHITFRMILESR---EKGEPSNCEGSKVYSHL 230

Db 182 VAQCEHIMETGKNRSVGYTLMMKDSRSHSITISMSAVDVG-----KDLHLAGKL 236

QY 231 NLVDLAGSRAAQTGAAGVRLKEGCNINRSLFILGVQVVKLSGQVGGFNYRDSKLTRI 290

Db 237 NLVDLAGSRSQKTGATGERLKEATKINLSLALGNVISALVDGRC-KHVPYRDSKLTRL 295

QY 291 LQNSLGNPKTRITCITPV--SPDETILALQFASTAKYMKNTPYNEVSTDEALLKRYR 348

Db 296 LQDSLGGNTKLMVACLSPADNNYDETLRLVANRANKINRKNRINEDDPKD-ALLREYQ 354

QY 349 KEIMDLKQLEEVSLERAQAMEKDQALQL-----BEKDLQKV----- 388

Db 355 ERIKKLKAII-----TQMSPSLSALLSRQVDPVQVEEKLPPQVIQHDMEAEK 406

QY 389 ---ONEKIENLRLMTVSSLTLOQLKAKRKRRVWCLGKINMKNSYNADQFNPTNI 445

Db 407 QLIREEYERLARL---KADYKAEQESRARLEEDIT-----AMRNSYDRLSTLEENL 456

QY 446 TTKTHK-LSINLL 457

Db 457 RKETEAVLQGVGL 469

RESULT 11

ID ADD49938 standard; protein; 1232 AA.

XX AC ADD49938;

XX DT 15-JAN-2004 (first entry)

XX DE Human lung specific tumour antigen L 1447p.

XX DE Human; lung cancer antigen; cytostatic; lung cancer; gene therapy;

XX KW vaccine; T-cell; tumour.

XX OS Homo sapiens.

XX OS US2003194764-A1.

XX PD 16-OCT-2003.

XX PF 04-APR-2002; 2002US-00116712.

XX PR 05-APR-2001; 2001US-0282289P.

XX PR 05-OCT-2001; 2001US-0327511P.

XX PA (CORI-) CORIXA CORP.

XX PA Bangor CS, Switzer A;

XX DR WPI; 2003-844452/78.

XX DR N-PSDB; ADD49936.

XX PT New isolated polypeptides and polynucleotides useful for diagnosing,

XX PT preventing and treating cancer, particularly lung cancer.

PS XX Claim 1; SEQ ID NO 670; 250pp; English.

CC The invention relates to an isolated polynucleotide (a) comprising any of

CC the 666 fully defined nucleotide sequences appearing as ADD49269 -

CC ADD49935, ADD49936 and ADD49938, complements of (a); sequences of at

CC least 20 contiguous residues of (a); sequences that hybridise to (a)

CC under highly stringent conditions; sequences having at least 75 or 90%

CC identity to (a); or degenerate variants of (a). Also included are an

CC isolated polypeptide (b) (comprising: sequences encoded by the new

CC polynucleotide; any of the 4 amino acid sequences fully defined in the

CC specification; or sequences having at least 70 or 90% identity to the

CC sequence in (a) or (b)), an expression vector comprising the above

CC polynucleotide operably linked to an expression control sequence, a host

CC cell transformed or transfected with the above expression vector, an

CC isolated antibody, or its antigen-binding fragment, that specifically

CC binds to the above polypeptide, an oligonucleotide that hybridises to the

CC above-mentioned nucleotide sequences under highly stringent conditions, a

CC fusion protein comprising at least one polypeptide cited above, detecting

CC the presence of a cancer in a patient (comprising: obtaining a biological

CC sample from the patient; contacting the biological sample with a binding

CC agent that binds to the polypeptide, or with the oligonucleotide cited

CC above; detecting in the sample an amount of the polypeptide that binds to

CC the binding agent, or an amount of a polynucleotide that hybridises to

CC the oligonucleotide; and comparing the amount of polypeptide, or

CC polynucleotide that hybridises to the oligonucleotide, to a predetermined

CC cut-off value and then determining the presence of a cancer in the

CC patient), a method for stimulating and/or expanding T-cells specific for

CC a tumour protein (comprising contacting T-cells with the above

CC polypeptide, polynucleotide or antigen-presenting cells that express the

CC polynucleotide, under conditions and for a time sufficient to permit the

CC stimulation and/or expansion of T-cells), an isolated T-cell population

CC comprising T-cells prepared by the method, a composition comprising a

CC first component selected from physiological carriers and

CC immunostimulants, and a second component selected from the above

CC polypeptide, polynucleotide, antibody, fusion protein, T-cell population

CC and antigen-presenting cells that express the above polypeptide,

CC stimulating an immune response in a patient (comprising administering to

CC the patient the above composition) treating lung cancer in a patient

CC (comprising administering to the patient the above composition and a

CC diagnostic kit (comprising: at least one oligonucleotide cited above; or

CC at least one antibody cited above and a detection reagent, where the

CC detection reagent comprises a reporter group). The composition and

CC methods are useful in diagnosing, preventing and treating cancer,

CC particularly lung cancer. The present sequence is a lung cancer-

CC associated antigen of the invention.

XX SQ

Sequence 1232 AA;

Query Match 29.6%; Score 697.5; DB 7; Length 1232;

Best Local Similarity 36.2%; Pred. No. 1.8e-47;

Matches 179; Conservative 79; Mismatches 178; Indels 59; Gaps 11;

QY 6 VAVCVVRPLNSREESLG-ETAQVYWKTDNNVIYQVDGSKSFNDRVFHGNETKNVYEE 64

Db 10 VRVALRCRPLVPKEISEGCGMCLSFVPEGPVVGTD--KSFTYDFVDFPSTEQEEVFNT 67

QY 65 IAAPIIDSAIQNGYNTIFAYGQTASGKTYTMMG-----SEDLGVIPRAIHDFQKIK 117

Db 68 AVAPLIKGVEKYNATVLAYGQTSGSKTYSMGGAATYAEQNEPTVGVIPIVQLLFKEID 127

QY 118 KFPDREFLLRVSYMEIYNETITDLCGTOKMKPLIREDVNRVNYVADLTEEVTYSEMA 177

Db 128 KKSDFEFTLVSYLEIYNEIIDLCPREKQAQINIREDPKGIKIVGLTEKTVLVALDT 187

QY 178 LKMITKGEKSRHYGETKMQORSRSHITFRMILESRKEGPFNSCEGSKVYSHLNLVDIAG 237

Db 188 VSCLEQGNNSRTVASTAMNSQSRSHAFTISLEQRKSD-----KNSFSRSLHLVDIAG 243

QY 238 SERAAQTGAAGVRLKEGCNINRSLFILGVQVVKLSGQVGGFNYRDSKLTRILONSLGG 297

Db 244 SEROKTKAEGDRLKEGININRGLLCGNVISALGDDKKGGFVPRDSKLTRLQDSLGG 303

QY 298 NPKTRITCITPV--SPDETILALQFASTAKYMKNTPYNEVSTDEALLKRYRKEIMDLK 355

Db 304 NSHTMTACVSPADSNLEETLTLRYADRAKIKKPKIVN-----IDPQTAELNHLK 355
 QY 356 KQLEEVSL-----ETRAQAMEKDQLAQLEEKDLQKQVQNEKIENL 396
 Db 356 QVQVQLVLLQAHGGTLPSTVPESENLSLMKQ--SLVEENKLSRGSEAAAGT 413
 QY 397 TRML---VTSSITLQQLKAKRRVTWCLGKINK-----MKNSYADQFNIPNTTKT 449
 Db 414 AQMLERIILTEQANEMKNAKLELRQHAACKLDLQKLVETLEQELKENVEIICNLQQLI 473
 QY 450 HKLSINLLREIDESV 464
 Db 474 TQLS-----DETV 481

RESULT 12

ADD18924

ID ADD18924 standard; protein; 1232 AA.

XX AC ADD18924;

XX DT 15-JAN-2004 (first entry)

XX DE Human disease related protein SeqID413.

XX KW human; disease state; cytostatic; antiinflammatory; ophthalmological;
 KW antiarteriosclerotic; vulnary; gene therapy;
 KW hypoxia-regulated condition; tumorigenesis; angiogenesis; apoptosis;
 KW inflammation; erythropoiesis; glycolysis; gluconeogenesis;
 KW glucose transportation; catecholamine synthesis; iron transport;
 KW nitric oxide synthesis; cancer; ischaemic condition; reperfusion injury;
 KW retinopathy; neonatal stress; pre-eclampsia; atherosclerosis;
 KW inflammatory condition; wound healing.

XX OS Homo sapiens.

XX PN WO2003018621-A2.

XX PD 06-MAR-2003.

XX PF 23-AUG-2002; 2002WO-GB003892.

XX PR 23-AUG-2001; 2001GB-00020558.

XX PR 05-OCT-2001; 2001GB-00024037.

XX PA (OXFO-) OXFORD BIOMEDICA UK LTD.

XX PI Kingsman SM, White J, Ward NR, Harris RA, Naylor S, Mundy CR;

XX DR WPI; 2003-290046/28.

XX DR N-PSDB; ADD18925.

XX PT New substantially purified polypeptide, useful for diagnosing or treating
 PT a hypoxia-regulated condition, such as cancer, ischemia, reperfusion
 PT injury, retinopathy, pre-eclampsia, atherosclerosis, inflammation, or
 PT wound healing.

XX PS Claim 25; SEQ ID NO 413; 424pp; English.

XX CC This invention relates to novel human genes and gene product which are
 CC implicated in certain disease states. Compounds which modulate the
 CC proteins of the invention may have cytostatic, antiinflammatory,
 CC ophthalmological, antiarteriosclerotic or vulnerary activities. The
 CC sequences of the invention may be useful for gene therapy. The invention
 CC may be useful for diagnosing or treating a hypoxia-regulated condition,
 CC such as tumorigenesis, angiogenesis, apoptosis, inflammation,
 CC erythropoiesis, or the biological response to hypoxia conditions
 CC including processes such as glycolysis, gluconeogenesis, glucose
 CC transportation, catecholamine synthesis, iron transport or nitric oxide
 CC synthesis. The disease includes cancer, ischaemic conditions, reperfusion
 CC injury, retinopathy, neonatal stress, pre-eclampsia, atherosclerosis,
 CC inflammatory conditions or wound healing. The present sequence is that of

CC a disease related protein of the invention.

XX SQ Sequence 1232 AA;
 Query Match 29.6%; Score 697.5; DB 7; Length 1232;
 Best Local Similarity 36.2%; Pred. No. 1.8e-47;
 Matches 179; Conservative 79; Mismatches 178; Indels 59; Gaps 11;

QY 6 VAVCVRVRLNSREESLG-ETAQVYWKTDNNVYIQVDGSKSFNPDVRFHGNKTNKYYEE 64
 Db 10 VRVALRCRLVPKSEISGQCMLSFVPGEPQVVGTD--KSFTYDFVDFDPSTQEVEFNT 67
 QY 65 IAAPIIDSLAQGYNGTIFAYGQTASGKTYTMG-----SEHGLGVIPRAIHDFQKIK 117
 Db 68 AVAPLKGKGVKGNVAVLAYGQTGSKTYSMGAYTAEQENETVGVIPRVIQLLFEKID 127
 QY 118 KFPDREFLLRVSYMEIYNETITDLLCGTQKMKPLIREDVNRNVYVADLTTEEVVYVTSEMA 177
 Db 128 KKSDFEFTLKVSYLEIYNEEILDLLCPSREKAQINREDPKGKIKIVGLTEKTVLVALDT 187
 QY 178 LKWITKGEKSRHYGETKMQQRSSRSHITFRMILESEKGEPSNCEGSKVSHLNLYDLAG 237
 Db 188 VSCLEQGNNSRTVASTAMNSQSSRSHAFITISLEQRKKS-----KNSSFRSKLHLVDLAG 243
 QY 238 SERAOTGAAGVRLKEGONINSLFILGOVVKKLSGQVGGFINYEDSKLTRILQNSLGG 297
 Db 244 SERQKTKAEGDRLEKGININGLCLGNVISALGDDKKGGFVYRDSKLTRILQNSLGG 303
 QY 298 NPKTRICTITPV--SFDETLTALOPASTAKYMKNTPYVNEVSTDEALLKRYRKEIMDLK 355
 Db 304 NSHTLMIACVSPADSNLEETLTLRYADRAKIKKPKIVN-----IDPQTAELNHLK 355
 QY 356 KQLEEVSL-----ETRAQAMEKDQLAQLEEKDLQKQVQNEKIENL 396
 Db 356 QVQVQLVLLQAHGGTLPSTVPESENLSLMKQ--SLVEENKLSRGSEAAAGT 413
 QY 397 TRML---VTSSITLQQLKAKRRVTWCLGKINK-----MKNSYADQFNIPNTTKT 449
 Db 414 AQMLERIILTEQANEMKNAKLELRQHAACKLDLQKLVETLEQELKENVEIICNLQQLI 473
 QY 450 HKLSINLLREIDESV 464
 Db 474 TQLS-----DETV 481

RESULT 13

ABG70992

ID ABG70992 standard; protein; 473 AA.

XX AC ABG70992;

XX DT 10-DEC-2002 (first entry)

XX DE Human target protein.

XX KW Microtubule stimulated ATPase; cellular proliferation; cancer; AIDS;
 KW hyperplasia; restenosis; cardiac hypertrophy; immune disorder; HIV;
 KW inflammation; autoimmune disease; arthritis; graft rejection; psoriasis;
 KW inflammatory bowel disease; proliferation; medical procedure; surgery;
 KW human immunodeficiency virus; acquired immunodeficiency syndrome;
 KW angioplasty; human; Hskif; kinesin family.

XX OS Homo sapiens.

XX PN US6440684-B1.

XX PD 27-AUG-2002.

XX PF 12-JUN-2000; 2000US-00592054.

XX PR 12-JUN-2000; 2000US-00592054.

XX PA (CYTO-) CYTOKINETICS INC.

ABG70991
ID ABG70991 standard; protein; 522 AA.
XX
AC ABG70991;
XX
DT 10-DEC-2002 (first entry)
XX
XX Human Hskif4 construct protein.
XX
XX Microtubule stimulated ATPase; cellular proliferation; cancer; AIDS;
XX hyperplasia; restenosis; cardiac hypertrophy; immune disorder; HIV;
XX inflammation; autoimmune disease; arthritis; graft rejection; psoriasis;
XX inflammatory bowel disease; proliferation; medical procedure; surgery;
XX human immunodeficiency virus; acquired immunodeficiency syndrome;
XX angioplasty; human; Hskif4; kinesin family.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH Region 1..22 /note= "N-terminal T7 epitope"
FT Region 496..516
FT Region /note= "C-terminal myc epitope"
FT Region 517..522
FT /note= "6-histidine residues at C-terminus"
XX US6440684-B1.
XX
XX 27-AUG-2002.
XX
XX 12-JUN-2000; 2000US-00592054.
XX
XX 12-JUN-2000; 2000US-00592054.
XX (CYTO-) CYTOKINETICS INC.
XX
XX Beraud C, Finer JT, Sakowicz R, Wood KW;
XX WPI; 2002-711529/77.
XX N-PSDB; ABS55161.
XX
XX Screening for modulators of target protein having microtubule stimulated
XX ATPase activity e.g. kinesin family of protein, useful for treating
XX cancer, psoriasis, arthritis, human immunodeficiency virus (HIV)
XX infection.
XX
XX Claim 2; Fig 4; 34pp; English.
XX
XX The present invention relates to a new method of screening modulators of
XX target protein with microtubule stimulated ATPase activity. The method
XX involves contacting the target protein with an agent at 1st and 2nd
XX concentrations and determining the level of activity (e.g. binding or
XX ATPase activity) of target protein, where a difference between levels of
XX activity of target protein contacted with 1st and 2nd concentrations of
XX an agent indicates that an agent modulates activity of target protein.
XX The invention can be used for screening for modulators of target protein
XX having microtubule stimulated ATPase activity. The compounds identified
XX by method of the invention are useful for treating cellular proliferation
XX including cancer, hyperplasia, restenosis, cardiac hypertrophy, immune
XX disorders and inflammation. The compounds identified by the method are
XX also useful for treating autoimmune disease, arthritis, graft rejection,
XX inflammatory bowel disease, proliferation induced by medical procedures,
XX e.g. surgery, angioplasty etc. The compounds are also useful for treating
XX psoriasis. The compounds are useful for inhibiting human immunodeficiency
XX virus (HIV) and thus treating acquired immunodeficiency syndrome (AIDS).
XX The present amino acid sequence represents the human target protein of
XX the invention
XX
XX Sequence 522 AA;
XX
XX Query Match 29.4%; Score 695; DB 5; Length 522;
XX Best Local Similarity 39.7%; Pred. No. 8.7e-48;
XX Matches 172; Conservative 70; Mismatches 155; Indels 36; Gaps 10;

XX Beraud C, Finer JT, Sakowicz R, Wood KW;
XX WPI; 2002-711529/77.
XX N-PSDB; ABS55162.
XX
XX Screening for modulators of target protein having microtubule stimulated
XX ATPase activity e.g. kinesin family of protein, useful for treating
XX cancer, psoriasis, arthritis, human immunodeficiency virus (HIV)
XX infection.
XX
XX Claim 3; Fig 6; 34pp; English.
XX
XX The present invention relates to a new method of screening modulators of
XX target protein with microtubule stimulated ATPase activity. The method
XX involves contacting the target protein with an agent at 1st and 2nd
XX concentrations and determining the level of activity (e.g. binding or
XX ATPase activity) of target protein, where a difference between levels of
XX activity of target protein contacted with 1st and 2nd concentrations of
XX an agent indicates that an agent modulates activity of target protein.
XX The invention can be used for screening for modulators of target protein
XX having microtubule stimulated ATPase activity. The compounds identified
XX by method of the invention are useful for treating cellular proliferation
XX including cancer, hyperplasia, restenosis, cardiac hypertrophy, immune
XX disorders and inflammation. The compounds identified by the method are
XX also useful for treating autoimmune disease, arthritis, graft rejection,
XX inflammatory bowel disease, proliferation induced by medical procedures,
XX e.g. surgery, angioplasty etc. The compounds are also useful for treating
XX psoriasis. The compounds are useful for inhibiting human immunodeficiency
XX virus (HIV) and thus treating acquired immunodeficiency syndrome (AIDS).
XX The present amino acid sequence represents the human target protein of
XX the invention
XX
XX Sequence 473 AA;
XX
XX Query Match 29.4%; Score 695; DB 5; Length 473;
XX Best Local Similarity 39.7%; Pred. No. 7.6e-48;
XX Matches 172; Conservative 70; Mismatches 155; Indels 36; Gaps 10;

QY 6 VAVCVVRPLNSRESIG-ETAYQVYKTDNNVIVQDGSKSFNDRVPHGNETTKNVEE 64
Db 8 VRVALRCRPLVPEKISECQCLSFVPGEPQVVGTD--KSFYDFVDPSTQEVEVNT 65
QY 65 IAAPTIDSAIGYNTIFAYGQTASGKTYTWG-----SEDHLGVIPRAIHDFQKIK 117
Db 66 AVAPLIKGVFGYNATVLAAGTSGSKTYSMGAYTAQEENPTVGVIPRVQLLKEID 125
QY 118 KFPDRFLLRYSMEIYNETITDLCGTQKMKPLIIRBVNVRNVVADLTREVVYSEMA 177
Db 126 KKSDFEFTLVKSYLBIYNEEILDLCPGREKAQINIREDPKEGKIVGITEKTVLVALDT 185
QY 178 LKWITKGRKSRHYGTQKQNSRSHHTIFRMLESREKGEPCSGVKVSHNLNVDLAG 237
Db 186 VSCLEQGNNSRTVASTAMNSQSSRSHAFITISLEQRKXSD-----KNSFRSKLHLVDLAG 241
QY 238 SRAACTGAAGVRLKEGGINNSLFTLGQVKKLSGQGVGFYNYRDSKLTILQNSLGG 297
Db 242 SERQKKTAEGRDLKEGINNKLCLGNVLSALGDDKGGFVYRDSKLTLLQDSLGG 301
QY 298 NPKTRITICTIPV--SFDETLALQFAPSTAKYMKNTPPYNEVSTDEALLKRYRKEIMDLK 355
Db 302 NSHTLMIACVSPADSNLEETLTLRYADRAKIKNKPVN-IDPQTAEHLNHLKQVY---- 356
QY 356 KQLEVSLETRAQM-----EKDQLAQLEBKDLLQKQVNEKIENTLMTLVSSSLTLQ 409
Db 357 QQLQVLLQAGHTLPGSITVPSFENLQSLMEKNQSLVENEK---LSRGLSEAAQQT--- 411
QY 410 QELKAKRKRRTVW 422
Db 412 ----AQMLERLIW 420

RESULT 14

Qy	6	VAVCVRVRPLNSRESLGG-ETAQVYWKTDNNVVIQVDGSGKSNFDFRVEHCHNETTKNVYEE	64
Db	30	VRVALRCRPLVPKEISEGCOMCLSPVGPQVVGTDD--KSFTYDFVFPDSTEQEEVFNT	87
Qy	65	IAAPIIDSAGIQYNGTIFAYGOTAGSKTYTMMG-----SEDHLGVIPIRAIHDFQKIK	117
Db	88	AVAPLIKGVFGYNATVLAYGOTGSGKTYSMGAVTAEQENEPVGVIPRVIQLLPKEID	147
Qy	118	KFPDREFLLRVSYMEIYNETITDLCGTCOMKPLIIREDDNRNVVADITEEVVYVSEMA	177
Db	148	KKSDFEFLKVSYLHLYNEEILDLCPREKAIQINREDPKGKIVGUTEKTVLVALDT	207
Qy	178	LKWITKGSKRHYGETKONQSRSSHTIFRMILEREKGEPCSGSVKSHNLVLDLAG	237
Db	208	VSCLEQGNNSRTVASTAMNSQSRSHAFTISLEQRKKSD----KNSFSRKLHLVDLAG	263
Qy	238	SERAAQCTAAGVRLKEGCNINRSLFILGOVIKLSDGQGVGFNTYRDSKLTIRLQNSLGG	297
Db	264	SEROKTKAEDRLKEGININRGLLCLGNVISALGDGDKGGVFPYRDSKLTIRLODSLGG	323
Qy	298	NPKTRIICTIPV--SPDETITALOFASTAKYMKNTPVYNEVSTDEALLKRYRKEIMDLK	355
Db	324	NSHTLMIACVSPADSNLEETLNTIYRDRARKIKNKPINW-IDPOTAEHLHKKQV---	378
Qy	356	KOLEEVSLETRAQAM-----EKOOLAOLLEEKDLLQKVONKIEINLFRMLVTSSSLTIQ	409
Db	379	QOLQVILLOAHGGHLPGSITVPEPSNLOSMLKQNSLVEENEK---LSRGLSEAAQOT--	433
Qy	410	QELKAKRKRRTVW	422
Db	434	-----AOMLERIIW	442

RESULT 15

AAM48337
ID AAM48337 standard; protein; 1038 AA.

DT 18-APR-2002 (first entry)

DE Human kinesin superfamily motor protein, KIF17.

KW Human; kinesin superfamily; motor protein; KIF17; microtubule binding;
KW nerve disease.

OS Homo sapiens.

PN JP2001333775-A.

04-DEC-2001.

24-MAY-2000; 2000JP-00153664.

PR 24-MAY-2000; 2000JP-00153664.

PA (SENT-) SENTAN KAGAKU GIJUTSU INCUBATION CENT KK.
PA (UYTY) UNIV TOKYO.

XX
DR WPI: 2002-135941/18

WFI; 2002-133941/18.
N-PSDB; ABA98475. ABA98176.

A kinesin superfamily motor protein KIF17, useful for the prevention of PT and/or treatment of nerve diseases caused by abnormal nerve cells.

Claim 1; Page 12-15; 23pp; Japanese.

The present sequence is human kinesin superfamily motor protein, KIF17. KIF17 is a microtubule binding motor protein which can be used for the prevention and/or treatment of nerve diseases caused by abnormal nerve cells

[illegible]

Search completed: July 29, 2004, 09:35:14
Job time : 28.9291 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 09:26:01 ; Search time 29.3136 Seconds
(without alignments)
4694.096 Million cell updates/sec

Title: US-10-045-631B-88_COPY_2_488

Perfect score: 2483

Sequence: 1 AEEGAVAVCVVRPLNSREE.....SDVFSNTLTLSEIENPAT 487

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04.*
1: Geneseq1980s.*
2: Geneseq1990s.*
3: Geneseq2000s.*
4: Geneseq2001s.*
5: Geneseq2002s.*
6: Geneseq2003as.*
7: Geneseq2003bs.*
8: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2483	100.0	2633	4	ABG06505 Novel hum
2	2483	100.0	2663	4	Aam39097 Human pol
3	2374.5	95.6	2688	4	Aam40883 Human pol
4	1584	63.8	2954	2	Aay01632 Amino aci
5	820	33.0	366	4	ABU53125 Intracell
6	781	31.5	348	4	ABU53208 Human cel
7	780.5	31.4	2013	4	ABBE62322 Drosophil
8	716.5	28.9	677	4	ABD49938 Human lun
9	713.5	28.7	1232	7	ADD18924 Human dis
10	710.5	28.6	1029	5	AAE17786 Human kin
11	710.5	28.6	1029	5	ADC10190 Human NOV
12	710.5	28.6	1029	5	ABG70993 Human Hsk
13	705.5	28.4	1234	5	ABG70993 Human Hsk
14	705	28.4	1034	3	Aag31112 Arabidops
15	705	28.4	1069	3	Aag31111 Arabidops
16	705	28.4	1121	3	Aag31110 Arabidops
17	703	28.3	1048	4	ABBS9245 Drosophil
18	703	28.3	1066	4	ABG67418 Amino aci
19	703	28.3	1066	6	ABG72693 Fruitfly
20	702.5	28.3	1038	6	Aam48337 Human kin
21	702.5	28.3	1232	7	ADD49937 Human kin
22	702.5	28.3	1232	7	ADD49932 Human lun
23	702	28.3	522	5	ABG70991 Human Hsk
24	696.5	28.1	1232	5	ABG70990 Human Hsk
25	695	28.0	473	5	ABG70992 Human tar

ALIGNMENTS

RESULT 1
ABG06505
ID ABG06505 standard; protein; 2633 AA.

XX AC ABG06505;

XX DT 13-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #6496.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX FN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US008631.

XX PR 31-MAR-2000; 2000US-00540217.

XX PR 23-AUG-2000; 2000US-00649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX N-PSDB; AAS70692.

XX New isolated polynucleotide and encoded polypeptides, useful in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits and to assess
biodiversity.

XX Claim 20; SEQ ID NO 36864; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
sequences. (I) is useful as hybridisation probes, polymerase chain
reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
and in recombinant production of (II). The polynucleotides are also used
in diagnostics as expressed sequence tags for identifying expressed
genes. (I) is useful in gene therapy techniques to restore normal
activity of (II) or to treat disease states involving (II). (II) is
useful for generating antibodies against it, detecting or quantitating a
polypeptide in tissue, as molecular weight markers and as a food
supplement. (II) and its binding partners are useful in medical imaging

Adb67093 Kinesin h
Aaw72746 Drosophil
Abb63485 Drosophil
Adb67088 Kinesin h
Aau19569 Human dia
Abp51294 Human MDD
Aag31117 Arabidops
Aag31116 Arabidops
Ade5349 Rat Prote
Aae14400 Human kin
Aau79590 Human kin
Abr48222 Human bla
Adb80468 Ovarian c
Adc35116 Human bre
Aam78880 Human pro
Adb67089 Kinesin h
Adb67091 Kinesin h
Aaw72745 Drosophil
Aaw72744 Drosophil
Aau74840 Human Hsk

26 664.5 26.8 1031 7 ADB67093
27 660.5 26.6 975 2 AAW72746
28 660.5 26.6 975 4 ABB63485
29 660.5 26.6 975 7 ADB67088
30 658 26.5 757 4 AAU19569
31 658 26.5 757 5 ABP51294
32 656.5 26.4 829 3 AAG31117
33 656.5 26.4 934 3 AAG31116
34 656 26.4 796 7 ADE5349
35 656 26.4 1388 5 AAE14400
36 656 26.4 1388 5 AAU79590
37 656 26.4 1388 6 ABR48222
38 656 26.4 1388 7 ADB80468
39 656 26.4 1388 7 ADC35116
40 654.5 26.4 963 4 AAM78880
41 654.5 26.4 963 7 ADB67089
42 653.5 26.3 963 7 ADB67091
43 652 26.3 411 2 AAW72745
44 652 26.3 441 2 AAW72744
45 652 26.3 1362 5 AAU74840

CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG3037 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directedly from WIPO at
CC ftp.wipo.int/pub/pubdirectly_fct_sequences
XX
SQ Sequence 2633 AA;

Query Match 100.0%; Score 2483; DB 4; Length 2633;
Best Local Similarity 100.0%; Pred. No. 7e-189;
Matches 487; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEEGAVAVCVVRPLNSREESLGETAQVYWKTDNNVIYQVDSKSFNDRVFGHNETTKN 60
DB 2 AEEGAVAVCVVRPLNSREESLGETAQVYWKTDNNVIYQVDSKSFNDRVFGHNETTKN 61
QY 61 VYEEIAAPIIDSAIQYNGTIFAYGQTASGKTYTMMGSEDLHGVIPRAIHDIQKIKKFP 120
DB 62 VYEEIAAPIIDSAIQYNGTIFAYGQTASGKTYTMMGSEDLHGVIPRAIHDIQKIKKFP 121
QY 121 DREFLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNVIYVADLTVEEVYTSEMAK 180
DB 122 DREFLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNVIYVADLTVEEVYTSEMAK 181
QY 181 ITKGEKSRHYGETQKQNSRSHTIFRMILESREKGEPSNCEGSKVSHNLNVDLAGSER 240
DB 182 ITKGEKSRHYGETQKQNSRSHTIFRMILESREKGEPSNCEGSKVSHNLNVDLAGSER 241
QY 241 AAOQTGAAGVRLKEGCNINRSFILGOVVKLSDQGVGFNYRDSKLTILQNSLGNPK 300
DB 242 AAOQTGAAGVRLKEGCNINRSFILGOVVKLSDQGVGFNYRDSKLTILQNSLGNPK 301
QY 301 TRIICITTPVSFDETLTALQASTAKYMNTPYNEVSTDEALKRYRKEIMDLKKOLEE 360
DB 302 TRIICITTPVSFDETLTALQASTAKYMNTPYNEVSTDEALKRYRKEIMDLKKOLEE 361
QY 361 VSLETRAQAMEKDQLAQLLEEKDLLQVQNEKIENLTMLVTSSSLTIQBELKAKRKRV 420
DB 362 VSLETRAQAMEKDQLAQLLEEKDLLQVQNEKIENLTMLVTSSSLTIQBELKAKRKRV 421
QY 421 TWCLGKINKMKNNSYADQFNPTNITTTKHLKSLINLLREIDSVCSDESDFNTLDTLSE 480
DB 422 TWCLGKINKMKNNSYADQFNPTNITTTKHLKSLINLLREIDSVCSDESDFNTLDTLSE 481
QY 481 IEMNPAT 487
DB 482 IEMNPAT 488

RESULT 2
AAM39097
ID AAM39097 standard; protein; 2663 AA.
XX
AC AAM39097;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 2242.
XX
XX Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX leukaemia.
XX
OS Homo sapiens.

XX WO200153312-A1.
XX
PD 26-JUL-2001.
XX
XX 26-DEC-2000; 2000WO-US034263.
XX
XX 23-DEC-1999; 99US-00471275.
XX 21-JAN-2000; 2000US-00488725.
XX 25-APR-2000; 2000US-00552317.
XX 20-JUN-2000; 2000US-00598042.
XX 19-JUL-2000; 2000US-00620312.
XX 03-AUG-2000; 2000US-00653450.
XX 14-SEP-2000; 2000US-00662191.
XX 19-OCT-2000; 2000US-00693036.
XX 29-NOV-2000; 2000US-00727344.
XX (HYSR-) HYSRQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX Wang J, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
XX Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI; 2001-442253/47.
XX N-PSDB; AAI58253.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders such
XX as central nervous system injuries.
XX
XX Example 4; SEQ ID NO 2242; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and the
XX encoded polypeptides (AAM38642-AAM42213) with nontropic,
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localised neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: Immune system suppression,
XX Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX assays for receptor activity, arthritis and inflammation, leukaemias and
XX C.N.S disorders. Note: The sequence data for this patent did not form
XX part of the printed specification

Sequence 2663 AA;

Query Match 100.0%; Score 2483; DB 4; Length 2663;
Best Local Similarity 100.0%; Pred. No. 7.2e-189;
Matches 487; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEEGAVAVCVVRPLNSREESLGETAQVYWKTDNNVIYQVDSKSFNDRVFGHNETTKN 60
DB 2 AEEGAVAVCVVRPLNSREESLGETAQVYWKTDNNVIYQVDSKSFNDRVFGHNETTKN 61
QY 61 VYEEIAAPIIDSAIQYNGTIFAYGQTASGKTYTMMGSEDLHGVIPRAIHDIQKIKKFP 120
DB 62 VYEEIAAPIIDSAIQYNGTIFAYGQTASGKTYTMMGSEDLHGVIPRAIHDIQKIKKFP 121
QY 121 DREFLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNVIYVADLTVEEVYTSEMAK 180
DB 122 DREFLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNVIYVADLTVEEVYTSEMAK 181
QY 181 ITKGEKSRHYGETQKQNSRSHTIFRMILESREKGEPSNCEGSKVSHNLNVDLAGSER 240
DB 182 ITKGEKSRHYGETQKQNSRSHTIFRMILESREKGEPSNCEGSKVSHNLNVDLAGSER 241
QY 241 AAOQTGAAGVRLKEGCNINRSFILGOVVKLSDQGVGFNYRDSKLTILQNSLGNPK 300
DB 242 AAOQTGAAGVRLKEGCNINRSFILGOVVKLSDQGVGFNYRDSKLTILQNSLGNPK 301

QY 301 TRIICITPVSFDETLTALQFASTAKYMKNTPVYNEVSTDEALLKRYRKEIMDLKKQLEE 360
 Db 302 TRIICITPVSFDETLTALQFASTAKYMKNTPVYNEVSTDEALLKRYRKEIMDLKKQLEE 361
 QY 361 VSLETRAQAMEKDQALAEKDLLOKQVNEKIENITRMLVTSSSLTLOQLKAKRKR 420
 Db 362 VSLETRAQAMEKDQALAEKDLLOKQVNEKIENITRMLVTSSSLTLOQLKAKRKR 421
 QY 421 TWCLGKINKMKNVADQFNIPNITTTKTKLSINLLREIDESVCSSEDFVSNLTDLSE 480
 Db 422 TWCLGKINKMKNVADQFNIPNITTTKTKLSINLLREIDESVCSSEDFVSNLTDLSE 481
 QY 481 IEWNPAT 487
 Db 482 IEWNPAT 488

RESULT 3
 AAM40883
 ID AAM40883 standard; protein; 2688 AA.
 AC AAM40883;
 XX
 DT 22-OCT-2001 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 5814.
 XX
 KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.
 XX
 OS Homo sapiens.
 XX
 FN WO200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000WO-US034263.
 XX
 PR 23-DEC-1999; 99US-00471275.
 PR 21-JAN-2000; 2000US-00488725.
 PR 25-APR-2000; 2000US-00552317.
 PR 20-JUN-2000; 2000US-00598042.
 PR 19-JUL-2000; 2000US-00620312.
 PR 03-AUG-2000; 2000US-00653450.
 PR 14-SEP-2000; 2000US-00662191.
 PR 19-OCT-2000; 2000US-00693036.
 PR 29-NOV-2000; 2000US-00727344.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
 PI Zhou P, Goodrich R, Drmanac RT;
 XX
 DR WPI: 2001-442253/47.
 DR N-PSDB; AAI60039.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders such
 PT as central nervous system injuries.
 XX
 PS Example 2; SEQ ID NO 5814; 10078pp; English.
 XX
 CC The invention relates to human nucleic acids (AAI57798-AAI61369) and the
 CC encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as

CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders. Note: The sequence data for this patent did not form
 CC part of the printed specification

XX Sequence 2688 AA;

Query Match 95.6%; Score 2374.5; DB 4; Length 2688;
 Best Local Similarity 96.7%; Pred. No. 3.5e-180;
 Matches 474; Conservative 2; Mismatches 11; Indels 3; Gaps 3;

QY 1 AEEGAVAVCVRVPLNSREESLGETAQVYWKTDNNVIYQVDSKSFNFDRVFGHNETTKN 60
 Db 23 AEEGAVAVCVRVPLNSREESLGETAQVYWKTDNNVIYQVDSKSFNFDRVFGHNETTKN 82
 QY 61 VYEEIAAPIIDSALIOGYNGTIFAYGQTASGKTYTWMGSEDLGLGVIPRA-IHDIP-OKIKK 118
 Db 83 VYEEIAAPIIDSALIOGYNGTIFAYGQTASGKTYTWMGSEDLGLGVIPQGFHGFQKIXE 142
 QY 119 -PPDREFLLRVSYMEIYNETITDLCGTCQMKPLIIRVDNVRNVVADLTEEVVYTSEWA 177
 Db 143 VPLDREFLLRVSYMEIYNETITDLCGTCQMKPLIIRVDNVRNVVADLTEEVVYTSEWA 202
 QY 178 LKWITGKSRHYGETKMNQSRSSHTIFRMILESRKGEPCNCGSVKVSHLNLVDLAG 237
 Db 203 LKWITGKSRHYGETKMNQSRSSHTIFRMILESRKGEPCNCGSVKVSHLNLVDLAG 262
 QY 238 SERAAQTGAAGVRLKEGNCINRSLEILGVQVKKLSDGVGGFINVRDSKLTILQNSLGG 297
 Db 263 SERAAQTGAAGVRLKEGNCINRSLEILGVQVKKLSDGVGGFINVRDSKLTILQNSLGG 322
 QY 298 NPKTRIICTITPVSFDETLTALQFASTAKYMKNTPVYNEVSTDEALLKRYRKEIMDLKKQ 357
 Db 323 NPKTRIICTITPVSFDETLTALQFASTAKYMKNTPVYNEVSTDEALLKRYRKEIMDLKKQ 382
 QY 358 LEEVSLETRAQAMEKDQALAEKDLLOKQVNEKIENITRMLVTSSSLTLOQLKAKRK 417
 Db 383 LEEVSLETRAQAMEKDQALAEKDLLOKQVNEKIENITRMLVTSSSLTLOQLKAKRK 442
 QY 418 RRVTCWLGKINKMKNVADQFNIPNITTTKTKLSINLLREIDESVCSSEDFVSNLTDLT 477
 Db 443 RRVTCWLGKINKMKNVADQFNIPNITTTKTKLSINLLREIDESVCSSEDFVSNLTDLT 502
 QY 478 LSEIEWNPAT 487
 Db 503 LSEIEWNPAT 512

RESULT 4

AAV01632
 ID AAV01632 standard; protein; 2954 AA.
 XX
 AC AAV01632;
 XX
 DT 22-JUN-1999 (first entry)
 XX
 DE Amino acid sequence of centromere-associated protein-E (CENP-E).

XX CENP-E; centromere-associated protein-E; ATPase activity;
 KW plus end-directed microtubule motor activity; chromosome congression;
 KW microtubule binding activity; chromosome movement; mitosis;
 KW cell proliferation; tumor; metastasis; vascular malfunction;
 KW inflammatory disease; immune disease; angiogenesis; hypertension;
 KW restenosis; fungal infection; selective herbicide; fungicide;
 KW insecticide; plant growth regulator; activator; cancer cell marker.
 XX
 OS Xenopus sp.
 XX
 FN W09913061-A1.

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 677 AA;

Query Match 28.9%; Score 716.5; DB 4; Length 677;
Best Local Similarity 40.8%; Pred. No. 2.3e-48; Indels 35; Gaps 13;
Matches 189; Conservative 73; Mismatches 166;

QY 6 VAVCVVRPLNSREESISGETAIVYKTDNNVIYQV-----DGSKSFNDRVFHGNETT 58
DB 21 VRVVVTRMDKNEISAGALSISVDKINRAITVMKPNATANEPPKTYIFDNVFDGGSNQ 80
QY 59 KNYVEIRAIPIIDSAIOGYNGTIFAYGQTASGKTYTMMGSEDH---LGVIPRAIHIFOK 115
DB 81 MDLYVDTARPIVDKVLGYNGTILAYGQTGKTYTMSGNPDSPQTKGIIPNAPAHIFGH 140
QY 116 IKFPPD-RBFLRVSVWEIYNETITDLCCTQKMKPLIREDVYNNVYVADLTVEEVYTS 174
DB 141 IAKAKENQKPLVRVSVWEIYNEEVRLD--GKOVGKSLEVKERPDIGVFVKDLGSGYVHNA 199
QY 175 EMALKWITGKESRHYGETKQNRSSRSHTIFRMILESREKGFSPNCEGSKVSHLNLVD 234
DB 200 DDLNEMILGNKRVAVGATQKQESSRSHAFITVERSELGEGD--VQVVRMGKQLQVD 257
QY 235 LAGSRAAQTGAAGVRLKEGCNINRSLFGLQVKKLSGQGVGFNYRDSKLTRELONS 294
DB 258 LAGSERQKTAQSGQRLKEATKINLSVLGNVISALVDGK--STHPIYRNSKLTRELQDS 316
QY 295 LGGNPKTRIICTITPV--SPDETILALQFASHTAKYMKNTPYVNEVSTDEALLKRYRKEIM 352
DB 317 LGGNSKTVMCATISPADSNYMETISTLYASRAKQIQRHMHINEEPKD-ALLRHFQBEIA 375
QY 353 DLKKOLEE-VSLETRAQAMEKQDLAQLLEEKDLQKQVNEKIENLTMLVTSLSLTQQ- 410
DB 376 RLRKQLEEGDSLEEPSSE-----EBEDTADDELEAPLE-----IELESSTIQAV 421
QY 411 ELKAKRRRVTC--LGKINKMKNVADQFNIPNTIKTKHL 452
DB 422 EKKPKKREKTDAREKEELAKRKEHKEIEHAKEQETLRNKL 464

RESULT 9
ADD49938
ID ADD49938 standard; protein; 1232 AA.
XX
AC ADD49938;
XX
DT 15-JAN-2004 (first entry)
XX
DE Human lung specific tumour antigen L 147p.
XX
KW Human; lung cancer antigen; cytostatic; lung cancer; gene therapy;
KW vaccine; T-cell; tumour.
XX
OS Homo sapiens.
XX
PN US2003194764-A1.
XX
PD 16-OCT-2003.
XX
PF 04-APR-2002; 2002US-00116712.
XX
PR 05-APR-2001; 2001US-0282289P.
PR 05-OCT-2001; 2001US-0327511P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Bangur CS, Switzer A;
XX

DB 61 DEMAKHIVHACQGFNGTIFAYGQTSSTGKTYTMMGDEQNPQVWVLAKEIFQOISSETER 120
QY 123 EPLLRVSVWEIYNETITDLCCTQKMKPLIREDVYNNVYVADLTVEEVYTSSEM-ALKWI 181
DB 121 DFLLRVGYEIEYNEKIYDLL--NKNQDLKIHESGNGIYVNVN--CBECIITSEVDLLRL 176
QY 182 TKGKSRHYGETKQNRSSRSHTIFRMILESREKGFSPNCEGSKVSHLNLVDLAGSERA 241
DB 177 CLGNKERTVGETNNERSRSHAFILIIESR-KSDHSD-DDAVIQSVLNLVDLAGSERA 234
QY 242 AQTGAAGVRLKEGCNINRSLFGLQVKKLSGQGVGFNYRDSKLTRELONS LGGNPKT 301
DB 235 DQTGARGARLKEGCHINKSLLFLSNVTKLSLENADNRFTNYRDSKLTRELQASLGNAFT 294
QY 302 RIICITTPVSFDITLALQFASHTAKYMKNTPYVNEVSTDEALLKRYRKEIMDLKKOLEEV 361
DB 295 SIITCTKPSIMESQSLSPATRAKKIRIKPQVNEWSDATMMKRLEREIKVLKDLAE- 353
QY 362 SLETRAQAMEKQDLAQLLEEKDLQKQVNEKIENLTSM-----LVTSSSLTLOQELKA 414
DB 354 -----EER---KQENQKVEHLERQIKHDMHKIICGHSLS---DKG 388
QY 415 KRKRRVTWC-----LGKINKMKNVADQF---NIPNTIKTKHL-----SI 454
DB 389 QKRRRTWCPTASGSHLELAETGTEDRIDQPKVSHLPKPVFFHTSNAGKRWDNIPKTI 448
QY 455 NILREIDESVCSRDVPS 472
DB 449 NILGSLD--IGTESNSIS 464

RESULT 8
ABB65183
ID ABB65183 standard; protein; 677 AA.
XX
AC ABB65183;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 22341.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR N-PSDB; ABL09286.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
PS Disclosure; SEQ ID NO 22341; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention

DR WPI: 2003-844452/78.
 XX N-PSDB; ADD49936.
 PT New isolated polypeptides and polynucleotides useful for diagnosing,
 PT preventing and treating cancer, particularly lung cancer.
 PS Claim 1; SEQ ID NO 670; 250pp; English.
 XX
 CC The invention relates to an isolated polynucleotide (a) comprising any of
 CC the 666 fully defined nucleotide sequences appearing as ADD49269 -
 CC ADD49936, ADD49936 and ADD49938, complements of (a); sequences of at
 CC least 20 contiguous residues of (a); sequences that hybridise to (a)
 CC under highly stringent conditions; sequences having at least 75 or 90%
 CC identity to (a); or degenerate variants of (a). Also included are an
 CC isolated polypeptide (b) (comprising: sequences encoded by the new
 CC polynucleotide; any of the 4 amino acid sequences fully defined in the
 CC sequence in (a) or (b)), an expression vector comprising the above
 CC polynucleotide operably linked to an expression control sequence, a host
 CC cell transformed or transfected with the above expression vector, an
 CC isolated antibody, or its antigen-binding fragment, that specifically
 CC binds to the above polypeptide, an oligonucleotide that hybridises to the
 CC above-mentioned nucleotide sequences under highly stringent conditions; a
 CC fusion protein comprising at least one polypeptide cited above, detecting
 CC the presence of a cancer in a patient (comprising: obtaining a biological
 CC sample from the patient; contacting the biological sample with a binding
 CC agent that binds to the polypeptide, or with the oligonucleotide cited
 CC above; detecting in the sample an amount of the polypeptide that binds to
 CC the binding agent, or an amount of a polynucleotide that hybridises to
 CC the oligonucleotide; and comparing the amount of polypeptide, or
 CC polynucleotide that hybridises to the oligonucleotide, to a predetermined
 CC cut-off value and then determining the presence of a cancer in the
 CC patient), a method for stimulating and/or expanding T-cells specific for
 CC a tumour protein (comprising contacting T-cells with the above
 CC polypeptide, polynucleotide or antigen-presenting cells that express the
 CC polynucleotide, under conditions and for a time sufficient to permit the
 CC stimulation and/or expansion of T-cells), an isolated T-cell population
 CC comprising T-cells prepared by the method, a composition comprising a
 CC first component selected from physiological carriers and
 CC immunostimulants, and a second component selected from the above
 CC polypeptide, polynucleotide, antibody, fusion protein, T-cell population
 CC and antigen-presenting cells that express the above polypeptide,
 CC stimulating an immune response in a patient (comprising administering to
 CC the patient the above composition) treating lung cancer in a patient
 CC (comprising administering to the patient the above composition and a
 CC diagnostic kit (comprising: at least one oligonucleotide cited above; or
 CC at least one antibody cited above and a detection reagent, where the
 CC detection reagent comprises a reporter group). The composition and
 CC methods are useful in diagnosing, preventing and treating cancer,
 CC particularly lung cancer. The present sequence is a lung cancer-
 CC associated antigen of the invention.
 XX
 SQ Sequence 1232 AA;
 Query Match 28.7%; Score 713.5; DB 7; Length 1232;
 Best Local Similarity 35.1%; Pred. No. 9.5e-48;
 Matches 182; Conservative 86; Mismatches 191; Indels 59; Gaps 11;
 QY 6 VAYCVRRPLNRSSEISG-ETAQVYWKTDNNVYQVDSKSENFDRVFHGNETKNVYEE 64
 DB 10 VRVALRCPLVPKSEISGCMQCLSFVGPQVVGTD--KSFYDFVDFDSTQEVEFNT 67
 QY 65 IAPITDSIAQVNGTIFAYGQTASGKTYTMMG-----SEHLGVIPRAIHDFOKIK 117
 DB 68 AVAPLKGVPKGNATVAYGQTGSGKTYSMGAYTAEQNEPTVGVPFVQILLKFEID 127
 QY 118 KFPDRFLLRVSMYELNETITLLCGTQKMKELIREDVNRNRYVADLTPEVYVYSEMA 177
 DB 128 KGSDFETLKVSYLETNEISLDLCPSEKQAQINREDPKGKIKVGLTEKTVLVALDT 187
 QY 178 LKWTIGKESRHYGKMNORSRSHITPFMILESREKGEPSNCEGSKVSHLVLDLAG 237
 DB 188 VSLCQGNNSRTVASTAMNSQSSRSHAITFISLEQRKSD----KNSFSRSLHLVDLAG 243

QY 238 SERAOTGAAGVRLKEGCMNINRSLFILQVVKLSDGQVGGFINYRDSKILTRILQNSLGG 297
 DB 244 SERQKTKAEGDRLEKGININRGLLCLGNVISALGDGKGGFVPRDSKILTRILQNSLGG 303
 QY 298 NPKTRIICTITPV--SFDLTLALQFASTAKYMKNTPYNEVSTDEALLKRYKEIMDLK 355
 DB 304 NSHTLMIACVSPADSNLEETLRLYADARKIKNPIVN-----IDPQTAEHLHLK 355
 QY 356 KOLEEVSLS-----ETRAQAMKDXQLAQLLEKXDLQKQVNEKIENL 396
 DB 356 QOVQQLQVLLQAHGGTLPSSITVPESENLSIMEKNQ--SLVEENEKLSRGSEAGQT 413
 QY 397 TWML---VTSSSLITLQOELKAKRKRVTWCLGKINK-----MKNSYADQFNPTNITTKT 449
 DB 414 AQMLERILTEQANEKMAKLELRQHAACKLDLQKLVELTDQELKENVEIICNLQQLI 473
 QY 450 HKLSINLLREIDESVCSDESDFNTLDTLSEIENPAT 487
 DB 474 TQLS-----DETACMAAIDTAVEQEAQVETSPET 504
 RESULT 10
 ADD18924
 ID ADD18924 standard; protein; 1232 AA.
 AC ADD18924;
 XX
 DT 15-JAN-2004 (first entry)
 DE Human disease related protein SeqID413.
 KW human; disease state; cytostatic; antiinflammatory; ophthalmological;
 KW antiarteriosclerotic; vulnary; gene therapy;
 KW hypoxia-regulated condition; tumorigenesis; angiogenesis; apoptosis;
 KW inflammation; erythropoiesis; glycolysis; gluconeogenesis;
 KW glucose transport; catecholamine synthesis;
 KW nitric oxide synthesis; cancer; ischaemic condition; iron transport;
 KW reinnopathy; neonatal stress; pre-eclampsia; atherosclerosis;
 KW inflammatory condition; wound healing.
 XX Homo sapiens.
 OS
 XX WO2003018621-A2.
 XX
 XX 06-MAR-2003.
 XX
 XX 23-AUG-2002; 2002WO-GB003892.
 XX
 XX 23-AUG-2001; 2001GB-00020558.
 XX
 XX 05-OCT-2001; 2001GB-00024037.
 XX
 XX (OXFO-) OXFORD BIOMEDICA UK LTD.
 XX
 XX Kingsman SM, White J, Ward NR, Harris RA, Naylor S, Mundy CR;
 XX
 XX WPI; 2003-230046/28.
 XX
 XX N-PSDB; ADD18925.
 XX
 PT New substantially purified polypeptide, useful for diagnosing or treating
 PT a hypoxia-regulated condition, such as cancer, ischemia, reperfusion
 PT injury, retinopathy, pre-eclampsia, atherosclerosis, inflammation, or
 XX wound healing.
 PS Claim 25; SEQ ID NO 413; 424pp; English.
 XX
 CC This invention relates to novel human genes and gene product which are
 CC implicated in certain disease states. Compounds which modulate the
 CC proteins of the invention may have cytostatic, antiinflammatory,
 CC ophthalmological, antiarteriosclerotic or vulnary activities. The
 CC sequences of the invention may be useful for gene therapy. The invention
 CC may be useful for diagnosing or treating a hypoxia-regulated condition,
 CC such as tumorigenesis, angiogenesis, apoptosis, inflammation,

CC erythropoiesis, or the biological response to hypoxia conditions
CC including processes such as glycolysis, gluconeogenesis, glucose
CC transportation, catecholamine synthesis, iron transport or nitric oxide
CC synthesis. The disease includes cancer, ischaemic conditions, reperfusion
CC injury, retinopathy, neonatal stress, pre-eclampsia, atherosclerosis,
CC inflammatory conditions or wound healing. The present sequence is that of
CC a disease related protein of the invention.
XX
SQ Sequence 1232 AA;

Query Match 28.7%; Score 713.5; DB 7; Length 1232;
Best Local Similarity 35.1%; Pred. No. 9.5e-48;
Matches 162; Conservative 86; Mismatches 191; Indels 59; Gaps 11;

QY 6 VAVCVVRPLNREESLG-ETAQVYWKTDNNVYVQDGSKSNFDFVHGNETTKNVEE 64
DB 10 VRVALRCPLVPKEISEGQMCLSFVPGEPQVVVGTD--KSFTYDFDPSTBEQEVFNT 67
QY 65 IAAPIIDSAIQVNGTIFAYGOTASGTYTMMG-----SEHLGVIPRAIHDFQKIK 117
DB 68 AVAPLIKGVKGNATVLAQVQSGKTYGWAQYTAQENETVGVIPRVIIQLLFKEID 127
QY 118 KPDPREFLLRVSYMEYNETITDLCCTQKMKPLIREDVNRNYYVADLTVEEVYTSMA 177
DB 128 KKSDFETLVSYLEIYNEIILDCPSREKAQINIREDPKEGKIIVGLTEKTVLVALDT 187
QY 178 LKWTGKESRHYGETKQNRQSRSHITFRMILESREKGPSCNCEGVKVSHLNVLVDLAG 237
DB 188 VSCLEQGNKRTVAATMNSQSSRSHAFTISLEQRKSD----KNSFSRKLHLVDLAG 243
QY 238 SERAAQTGAAGVRLKEGNNINRSFILGOVVKLSDGVGGINYRDSKLTIRILQNSLGG 297
DB 244 SERQKTKAEGDRLEKGININRGLLCLGNVISALGDDKGGFVYRDSKLTIRILQNSLGG 303
QY 298 NPKTRIITITPV--SPDETITLQFASAKYMNQTPYVNEVSTDEALLKRYKEIMDLK 355
DB 304 NSHTLMACVSPADSNLEETLTLRYADRAKIKNKPVN-----IDPQTAEHLNHLK 355
QY 356 KQLEEVSL-----ETRAQAMEKDLQALBEKDLQKQVQEKIENL 396
DB 356 QOVQOLQVLLQAHGGTLPGSITVPEPSNLSQSLMEKNQ--SLVSENEKLSGLSEAAQOT 413
QY 397 TRML---VTSSSLTLQELKAKRRVTVCLGKINK---MKNSNYADQFNIPNTITKT 449
DB 414 AQMLERILITQANEKMAKLEELRQHAACKLDLQKLVEITLEDQELKENVEILCNLQOLI 473
QY 450 HKLSINLLREIDESVCSSESDVFSNTLDTLSBIENNPAT 487
DB 474 TQLS-----DETVACMAAIDTAVEQEAQVETSPET 504

RESULT 11
AAE17786
ID AAE17786 standard; protein; 1029 AA.
XX
AC AAE17786;
XX
DT 07-MAY-2002 (first entry)
XX Human kinesin superfamily motor protein, HsKif17.
DE
XX Human; kinesin superfamily motor protein; HsKif17; autoimmune disease;
KW cellular proliferation; therapy; cancer; hyperplasia; graft rejection;
KW cardiac hypertrophy; immune disorder; inflammation; immunosuppressive;
KW angiodysplasia; arthritis; restenosis; inflammatory bowel disease;
KW cytostatic; vasotropic; antiarthritic; antiinflammatory.
XX
OS Homo sapiens.
XX WO200198314-A2.
XX
XX
PD 27-DEC-2001.
XX

PF 20-JUN-2001; 2001WO-US019811.
XX
PR 20-JUN-2000; 2000US-00597602.
XX
XX (CYTO-) CYTOKINETICS INC.
PA
XX Beraud C, Freedman R;
PI
DR WPI; 2002-147789/19.
DR N-PSDB; AAD28568.
XX
XX Novel human kinesin superfamily motor protein, HsKif17, useful for
PT identifying modulators that may be used to treat cell proliferative
PT disorders, e.g., cancer, hyperplasias, restenosis, autoimmune disease,
PT arthritis, graft rejection.
XX
PS Claim 11; Fig 2; 64pp; English.

XX The present invention relates to an isolated novel human kinesin
CC superfamily motor protein, HsKif17. HsKif17 is useful for screening
CC modulators of HsKif17, comprising contacting biologically active HsKif17
CC with a candidate agent in a test and control concentration and assaying
CC for the level of HsKif17 activity, where a change in activity between the
CC test and control concentration indicates a modulator, and where screening
CC occurs in a multiwell plate as a part of high-throughput screen. The
CC modulators identified are useful in treatment of cellular proliferation
CC diseases such as cancer, hyperplasias, restenosis, cardiac hypertrophy,
CC immune disorders and inflammation, e.g., autoimmune disease, arthritis,
CC graft rejection, inflammatory bowel disease, proliferation induced after
CC medical procedures such as surgery, angioplasty, etc. The present
CC sequence is human HsKif17

XX Sequence 1029 AA;
Query Match 28.6%; Score 710.5; DB 5; Length 1029;
Best Local Similarity 36.5%; Pred. No. 1.3e-47;
Matches 190; Conservative 80; Mismatches 176; Indels 75; Gaps 16;

QY 5 AVAVCVVRPLNREESLGETAQVYWKTD-----NNVYVQDGSKSNFDFVHGNE 56
DB 5 AVKVVVRCRPMNQRELE--RCQPVVTVDCARAOCCIQNGAEDPQKQFFDGAHYVDH 62
QY 57 TTKNVYVEIAPIIDSAIQVNGTIFAYGOTASGTYTMMGSED---HLGVIPRAIHDF 113
DB 63 VTEQIYNEIAVPLVEGTEGNGTIFAYGQSGSKFTMQGLDPPSPQRGIIPRAFEHVF 122
QY 114 QKIKKFPDPREFLLRVSYMEYNETITDLCCTQKMKPLIREDVNRNYYVADLTVEEVY 173
DB 123 ESVQCAENTKFLVRASYLEIYNEIDVRLDGLADTKQK-LLEKHEPEKGVYVKGSLMHTVHS 181
QY 174 SEMALKWITKGEKSRHYGETKQNRQSRSHITFRMILES---EKGEPSNCEGVSKVSHL 230
DB 182 VAQCEHIMETGKQNRVSGVITLNNKDSRSHSIFITISMSAVDERG----KHLRAGKL 236
QY 231 NLVDLAGSERAQAGVRLKEGNNINRSFILGOVVKLSDGVGGINYRDSKLTIRI 290
DB 237 NLVDLAGSERSQKGTGTERLKEATKINLSALGNVISALVDGRC-KHVPYRDSKLTIRL 295
QY 291 LQNSLGGNPKTRIITITPV--SPDETITLQFASAKYMNQTPYVNEVSTDEALLKRYR 348
DB 296 LQDSLGGNKTLLWACLSPADNNYDETILSTLYANRAKRNKRNKRNEDPKD-ALLREYQ 354
QY 349 KEIMDLKQLEEVSLVETRAQAMEKDLQAL-----EEKDLQKV----- 388
DB 355 EBIKKKAIL-----TQQMSPSLSALLSRQVPPVQVVEEKLPPQVQIHDMEAEK 406
QY 389 ---QNEKTEINLRMLVTSLSLTLQOELKAKRRVTVCLGKINKMKNKSNYADQFNIP 445
DB 407 QLIREYERLARL---KADYKAEQESRARLEEDIT-----AMENSYDVRSLTEENL 456
QY 446 TTKTKHKLINLLREIDESVCSSESDVFSNTLDTLSBIENNP 486
DB 457 RKETEAV-----LQGVGLYKAEVMSRA-EFASAEYPPA 489

RESULT 12
ADCL0190
ID ADC10190 standard; protein; 1029 AA.
AC ADC10190;
XX
DT 18-DEC-2003 (first entry)
XX
DE Human NOVX polypeptide SEQ ID NO: 212.
XX
KW cytostatic; antidiabetic; anorectic; cerebroprotective; neuroprotective;
KW antiinflammatory; gene therapy; antisense therapy; thyromimetic; NOVX;
KW pathology; cancer; diabetes; obesity; endocrine disorder; CNS disorder;
KW inflammatory disorder; chromosome mapping; tissue typing;
KW predictive medicine.
XX
OS Homo sapiens.
XX
PN WO2003000842-A2.
XX
PD 03-JAN-2003.
XX
PF 04-JUN-2002; 2002WO-US017443.
XX
PR 04-JUN-2001; 2001US-0295607P.
PR 04-JUN-2001; 2001US-0295661P.
PR 06-JUN-2001; 2001US-0296404P.
PR 06-JUN-2001; 2001US-0296418P.
PR 07-JUN-2001; 2001US-0296575P.
PR 11-JUN-2001; 2001US-0297414P.
PR 12-JUN-2001; 2001US-0295573P.
PR 12-JUN-2001; 2001US-0297567P.
PR 14-JUN-2001; 2001US-0298285P.
PR 15-JUN-2001; 2001US-0298528P.
PR 18-JUN-2001; 2001US-0299133P.
PR 19-JUN-2001; 2001US-0299230P.
PR 21-JUN-2001; 2001US-0299949P.
PR 22-JUN-2001; 2001US-0300177P.
PR 26-JUN-2001; 2001US-0300883P.
PR 28-JUN-2001; 2001US-0301530P.
PR 28-JUN-2001; 2001US-0301550P.
PR 03-JUL-2001; 2001US-0302951P.
PR 31-JUL-2001; 2001US-0308909P.
PR 14-SEP-2001; 2001US-0322297P.
PR 25-SEP-2001; 2001US-0324669P.
PR 03-DEC-2001; 2001US-0337477P.
PR 14-DEC-2001; 2001US-0341562P.
PR 21-FEB-2002; 2002US-0358565P.
PR 21-FEB-2002; 2002US-0359122P.
PR 22-FEB-2002; 2002US-0358978P.
PR 22-FEB-2002; 2002US-0359034P.
PR 22-FEB-2002; 2002US-0359035P.
PR 27-FEB-2002; 2002US-0359121P.
PR 27-FEB-2002; 2002US-0359964P.
PR 01-MAR-2002; 2002US-0360859P.
PR 12-MAR-2002; 2002US-0363430P.
PR 12-MAR-2002; 2002US-0363678P.
PR 10-APR-2002; 2002US-0371346P.
PR 10-MAY-2002; 2002US-0379444P.
PR 04-JUN-2002; 2002US-00379444.
XX
(CURA-) CURAGEN CORP.
XX
PI Agee ML, Anderson DW, Berghs C, Casman SJ, Catterton E;
PI Dipippo VA, Edinger SR, Eisen A, Ellerman K, Gangolli EA;
PI Gerlach VL, Gorman I, Guo X, Herrmann JL, Hjal T, Ji W, Kekuda R;
PI Khramtsov NV, Li L, Liu X, Malyankar UM, Miller CE, Millet I;
PI Ort T, Padigar M, Patturajan M, Pena CEA, Rastelli L, Rieger DK;
PI Rothenberg ME, Shency SG, Shimkets RA, Smithson G, Spaderna SK;
PI Spytek KA, Stone DJ, Vernet CAM, Zhong H, Zhong M, Alsobrook JP;
PI Burgess CE, Lepley DM;

XX WPI: 2003-210149/20.
DR N-PSDB; ADC10189.
XX
PT New isolated NOVX polypeptides and nucleic acid molecules useful for
PT treating, preventing and diagnosing pathological conditions with NOVX-
PT associated disorders, such as cancer, obesity, diabetes and inflammatory
PT or CNS diseases.
XX
PS Claim 1; SEQ ID NO 212; 772pp; English.
XX
CC The invention relates to novel isolated polypeptides, mature form of the
CC polypeptide, a sequence that is 95% identical to the polypeptide or the
CC polypeptide comprising one or more conservative substitutions. The NOVX
CC polypeptide is useful for treating or preventing a pathology associated
CC with the polypeptide e.g. disorders associated with aberrant expression
CC or activity of the polypeptide, such as cancer, diabetes, obesity, and
CC endocrine, CNS and inflammatory disorders. They can also be used in
CC various detection and screening assays, chromosome mapping, tissue typing
CC and predictive medicine. This sequence corresponds to one of the
CC polypeptides of the invention.
XX
SQ Sequence 1029 AA;
Query Match 28.6%; Score 710.5; DB 7; Length 1029;
Best Local Similarity 36.5%; Pred. No. 1.3e-47;
Matches 190; Conservative 80; Mismatches 176; Indels 75; Gaps 16;
QY 5 AVAVCVRRVPLNSREESLGETAQVYWKTD-----NNVIYQVDSKSFNDRVPHNE 56
Db 5 AVKVVRRCRPMQRBREL--RCQPVTVVDCARAQCQIQNGAADPEPKQTFDGAHVHDH 62
QY 57 TTKNVEHIAPIIDSAITQGYNGTIFAYGQTASGKTYTMGSED---HLGVIPRAIHDI 113
Db 63 VTEQIYNEIAYPLVEGVTGYNGTIFAYGQTGSGKFTMQGLPDPSPQRGIIPRAFEHVF 122
QY 114 QKIKKPPDFRFLRVSYMBIYNETITDLCGQKMKPLIREDVNRNVVADLTBRVVVT 173
Db 123 ESVQCAENTKFLVRASYLEIYNEVDRLDGLGATKQK-LELKEHPEKGVTVKGLSMITVHS 181
QY 174 SEMALKWITKGEKSRHYGETKQNRSSRSHTIFRMLESR---EKGEPSNCSGVSKVSHL 230
Db 182 VAQCEHIMETGWNRSVGVTLNKNDSRSHTSFTSIEMSAVDERG-----KDLRAGKL 236
QY 231 NLVDLAGSERRAQTCAAGVRLKEGNCINRSLFILQVVKLSDGQGVGFINTRDSKLTPI 290
Db 237 NLVDLAGSERSQKTGATGERLKEATKINLSLALGNVISALVDGRC-KHVPYRDSKLTPL 295
QY 291 LQNSLGNPKTRIICITIPV--SPDETLTALQFASTAKYMKNTPVVNEVSTDEALLKRYR 348
Db 296 LQDSLGGNTKILMVACLSPADNNYDETSLTRYANRKNRKNKPRINEDPKD-ALLREYQ 354
QY 349 KEIMDLKKOLEVSLERAQAMEKDQLAQL-----EEKDLLQKV-----388
Db 355 BEIKKILKAIL-----TQMSPSLSALLSRQVPPDPVQVEEKLPPQVLIQHDMEAK 406
QY 389 ---QNEKJENITRLMTVSSSLTLOELKAKKRRVTVCLGKINKMKNYADQFNIPNI 445
Db 407 QLIREEVEERLARL---KADYKAEQESRLEEDIT-----AMRNSYDVRSLTEENL 456
QY 446 TTKTKHLSINLREIDESVCSSESVFNTLDTLSIENWPA 486
Db 457 RKETEAV-----LQGVLYKAEVMSRA-EFASAEYPPA 489
RESULT 13
ABG70993
ID ABG70993 standard; protein; 1234 AA.
XX
AC ABG70993;
XX
DT 10-DEC-2002 (first entry)
XX

PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136332P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159291P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.

PR	22-SEP-1999;	99US-0155139P;
PR	23-SEP-1999;	99US-0155486P;
PR	24-SEP-1999;	99US-0155659P;
PR	28-SEP-1999;	99US-0156458P;
PR	29-SEP-1999;	99US-0156569P;
PR	04-OCT-1999;	99US-0157117P;
PR	05-OCT-1999;	99US-0157753P;
PR	06-OCT-1999;	99US-0157865P;
PR	07-OCT-1999;	99US-0158029P;
PR	08-OCT-1999;	99US-0158232P;
PR	12-OCT-1999;	99US-0158369P;
PR	13-OCT-1999;	99US-0159233P;
PR	13-OCT-1999;	99US-0159294P;
PR	14-OCT-1999;	99US-0159295P;
PR	14-OCT-1999;	99US-0159329P;
PR	14-OCT-1999;	99US-0159330P;
PR	14-OCT-1999;	99US-0159331P;
PR	14-OCT-1999;	99US-0159637P;
PR	14-OCT-1999;	99US-0159638P;
PR	18-OCT-1999;	99US-0159584P;
PR	21-OCT-1999;	99US-0160741P;
PR	21-OCT-1999;	99US-0160747P;
PR	21-OCT-1999;	99US-0160767P;
PR	21-OCT-1999;	99US-0160768P;
PR	21-OCT-1999;	99US-0160770P;
PR	21-OCT-1999;	99US-0160814P;
PR	21-OCT-1999;	99US-0160815P;
PR	22-OCT-1999;	99US-0160980P;
PR	22-OCT-1999;	99US-0160981P;
PR	22-OCT-1999;	99US-0160989P;
PR	25-OCT-1999;	99US-0161404P;
PR	25-OCT-1999;	99US-0161405P;
PR	25-OCT-1999;	99US-0161406P;
PR	26-OCT-1999;	99US-0161359P;
PR	26-OCT-1999;	99US-0161360P;
PR	26-OCT-1999;	99US-0161361P;
PR	28-OCT-1999;	99US-0161920P;
PR	28-OCT-1999;	99US-0161922P;
PR	28-OCT-1999;	99US-0161993P;
PR	29-OCT-1999;	99US-0161242P;

Best Local Similarity 34.3%; Pred. No. 3.7e-47;
Matches 196; Conservative 96; Mismatches 148; Indels 132; Gaps 18;

```

Qy 1 AEGAVAVCVVRPLN-----SRBES-----LGETA---QVWKTDN 34
Db      : : : : | | | | : : : : | | | | : : : : | | | |
Qy 42 SERDSISVVRPEPLRYARSLANDQIYCRYSHVDAGVNSLLGYQRGDEVAVTPDG 101
Db      : : : : | | | | : : : : | | | | : : : : | | | |
Qy 35 NVI--YQVDSGSFNDRVFHGNETTKNVYBEIAAPIDSIAIOGNYGTIFAYGOTASGKT 92
Db      : : : : | | | | : : : : | | | | : : : : | | | |
Qy 102 DTLVRHEYNPLTAYAFDKVFGPQAATTIDYDVAARPVVKAAMEGVNGTVFAYGYTSGKT 151
Db      : : : : | | | | : : : : | | | | : : : : | | | |
Qy 93 YTM-----MGSEDLHGVIPIRAIHDF 113
Db      : : : : | | | | : : : : | | | | : : : : | | | |
Qy 162 HTMHVRVLKKNKPYTFPSILVFLQNTCAWMDLFFNLVPLWYLLGDQESPGLIPLAKDVF 221
Db      : : : : | | | | : : : : | | | | : : : : | | | |
Qy 114 QXIKF-----PDREFLLRYSYWEIYNETTDLGCGTQKMKPLIIBEDVNRVNVYADLT 167
Db      : : : : | | | | : : : : | | | | : : : : | | | |
Qy 222 STIQDVSGLNGTPGREFLLRSYLEIYNEVINDLLDPTG--QNLVRVED--SQGYVEGK 278
Db      : : : : | | | | : : : : | | | | : : : : | | | |
Qy 168 EEVWYTSWALKWITKGKSRHYGETKQKQBSRSHTFIRMILESREKGFSPSCGSGKV 227
Db      : : : : | | | | : : : : | | | | : : : : | | | |
Qy 279 EEVLSPGHALSFIAAGEEHRHVGSNNFLNLSRSSHTTFTLMVSSATGDEYD---GVIF 335
Db      : : : : | | | | : : : : | | | | : : : : | | | |
Qy 228 SHLNLDVLAGSERAAQTGAAGVRUKEGECNINRSILFGQVVKLSDQVGFGFNYPDRSKL 287
Db      : : : : | | | | : : : : | | | | : : : : | | | |
Qy 336 SQNLNLDLAGSE--SKYETTGLRKEGYSINKSLTLGTVTIGKLGSEK--ATHIYPDRSKL 393
Db      : : : : | | | | : : : : | | | | : : : : | | | |
Qy 288 TRILQNSLGGNPKTRIICTITPV--SPDETALQFASTAKYMKNTPVYNEVSTDEALLK 345
Db      : : : : | | | | : : : : | | | | : : : : | | | |
Qy 394 IELLQSSLSHGHSVSLICTITPASSSSEETHNTLKFASRAKSTHIVASRNQIIDKSLIK 453
Db      : : : : | | | | : : : : | | | | : : : : | | | |
Qy 346 RYRKEIMDLKOLBEV-----SLETRAQAMEKQD--LQAELLEKDLLOKVONEK 392

```

Fri Aug 6 10:49:12 2004

Db	454	KYOREISTKLELDQRRGMLVGVSHHELSLKQQLBECQVWQSRLEBEEEAALMSR	513
QY	393	IENLTRLMLVTSSSLTIQOELKAKRRRVTWCLGKINKMKNSNYADOENIPTNITTKHL	452
Db	514	IQKLTKLILVST-----KNSIPGYSGDIP-----THQR	541
QY	453	SINLREID-ESVCSSESVF---SNTLDTLSE	480
Db	542	SLSAGKDDKFDLSLLESNDLGSPTSLLISE	573

Search completed: July 29, 2004, 09:35:15
Job time : 30.3136 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 09:30:02 ; Search time 8.62775 Seconds

(without alignments)
2914.068 Million cell updates/sec

Title: US-10-045-631B-88_COPY_2_488

Perfect score: 2483

Sequence: 1 AEGAVAVCVVRPLNSRRE.....SDVFSNTLDTLSIEWNPAT 487

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/2/iaa/5A COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2483	100.0	2662	4	US-09-595-684B-31
2	1584	63.8	2954	4	US-09-150-867-1
3	705.5	28.4	1234	4	US-09-592-054-8
4	704	28.4	1231	4	US-09-595-684B-23
5	703	28.3	1066	3	US-09-541-782-8
6	703	28.3	1066	4	US-09-723-820-8
7	703	28.3	1066	4	US-10-270-085-8
8	702	28.3	522	4	US-09-592-054-4
9	696.5	28.1	1232	4	US-09-592-054-2
10	695	28.0	473	4	US-09-592-054-6
11	664.5	26.8	1031	4	US-09-914-259-24
12	660.5	26.6	975	4	US-09-914-259-19
13	656	26.4	1388	4	US-09-572-191-2
14	656	26.4	1388	4	US-09-723-262-2
15	656	26.4	1388	4	US-09-723-219-2
16	654.5	26.3	963	4	US-09-914-259-20
17	653.5	26.3	963	4	US-09-914-259-22
18	652	26.3	411	2	US-08-713-815A-4
19	652	26.3	441	2	US-08-713-815A-3
20	642.5	25.9	967	4	US-09-914-259-21
21	640	25.8	957	4	US-09-914-259-16
22	638.5	25.7	1032	4	US-09-914-259-26
23	636.5	25.6	956	4	US-09-914-259-17
24	635.5	25.6	1027	4	US-09-914-259-27
25	632.5	25.5	1056	4	US-09-595-684B-29
26	632.5	25.5	1057	4	US-09-541-782-10
27	632.5	25.5	1057	4	US-09-723-820-10

28	632.5	25.5	1057	4	US-10-270-085-10	Sequence 10, Appl
29	632	25.5	409	4	US-09-572-191-6	Sequence 6, Appl
30	632	25.5	409	4	US-09-723-262-6	Sequence 6, Appl
31	632	25.5	409	4	US-09-723-219-6	Sequence 6, Appl
32	628	25.3	815	4	US-09-914-259-18	Sequence 18, Appl
33	624.5	25.2	575	4	US-09-724-519-8	Sequence 8, Appl
34	624.5	25.2	575	4	US-09-592-037-8	Sequence 8, Appl
35	624.5	25.2	575	4	US-09-428-156B-8	Sequence 8, Appl
36	623.5	25.1	1057	4	US-09-428-156B-8	Sequence 8, Appl
37	621.5	25.0	935	4	US-09-914-259-25	Sequence 25, Appl
38	613.5	24.7	513	4	US-09-724-519-6	Sequence 6, Appl
39	613.5	24.7	513	4	US-09-592-037-6	Sequence 6, Appl
40	613.5	24.7	513	4	US-09-428-156B-6	Sequence 6, Appl
41	611.5	24.6	1111	4	US-09-914-259-28	Sequence 28, Appl
42	610	24.6	928	4	US-09-914-259-23	Sequence 23, Appl
43	608	24.5	1053	4	US-09-724-519-2	Sequence 2, Appl
44	608	24.5	1053	4	US-09-592-037-2	Sequence 2, Appl
45	603	24.3	375	4	US-09-572-191-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1

US-09-595-684B-31

; Sequence 31, Application US/09595684B

; Patent No. 6544766

; GENERAL INFORMATION:

; APPLICANT: Beraud, Christophe

; APPLICANT: Ohashi, Cara

; APPLICANT: Sakowicz, Roman

; APPLICANT: Vaisberg, Eugeni

; APPLICANT: Wood, Kenneth

; APPLICANT: Yu, Ming

; TITLE OF INVENTION: Human kinesins and methods of producing

; TITLE OF INVENTION: and purifying human kinesins

; FILE REFERENCE: cytop036

; CURRENT APPLICATION NUMBER: US/09/595,684B

; CURRENT FILING DATE: 2002-06-24

; PRIOR APPLICATION NUMBER: 09/295,612

; PRIOR FILING DATE: 2000-04-20

; NUMBER OF SEQ ID NOS: 105

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 31

; LENGTH: 2662

; TYPE: PRT

; ORGANISM: Human

US-09-595-684B-31

Query Match

Best Local Similarity 100.0%; Score 2483; DB 4; Length 2662;

Mismatches 487; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	AEGAVAVCVVRPLNSRREISGETAQVYWKTDNNVIYVDGSKSFNDFVPHGNETTKN	60
DB	2	AEGAVAVCVVRPLNSRREISGETAQVYWKTDNNVIYVDGSKSFNDFVPHGNETTKN	61
QY	61	VVEETAAPIIDSAIQYNGTIFAYGTASGKTYTWMGSEDLGVIPRAIHDFQKIKKFP	120
DB	62	VVEETAAPIIDSAIQYNGTIFAYGTASGKTYTWMGSEDLGVIPRAIHDFQKIKKFP	121
QY	121	DREFFLLRYSYMEIYNETITDLCGQKMKPLIREDVRNVYVADLTERVYVTSMALKW	180
DB	122	DREFFLLRYSYMEIYNETITDLCGQKMKPLIREDVRNVYVADLTERVYVTSMALKW	181
QY	181	ITKGEKSRHYGTQKNQSSRSHTIFRMLESREKGEPSNCEGSKVSHNLVDLAGSR	240
DB	182	ITKGEKSRHYGTQKNQSSRSHTIFRMLESREKGEPSNCEGSKVSHNLVDLAGSR	241
QY	241	AAQTGAAGVRLKEGNCIINRSFILQVVKLSLSDGQVGGFINYRDSKLTILQNSLGGNPK	300
DB	242	AAQTGAAGVRLKEGNCIINRSFILQVVKLSLSDGQVGGFINYRDSKLTILQNSLGGNPK	301

```

QY 301 TRIICITTPVSFDETLTALQFASAKYMKNTPPYNEVSTDEALKRYRKEIMDLKKQLEE 360
Db 302 TRIICITTPVSFDETLTALQFASAKYMKNTPPYNEVSTDEALKRYRKEIMDLKKQLEE 361
QY 361 VSLTRAQAMEKQDLAQLLEKDLQVQNEKINLRLMVLTSSTLQOELKAKRRRV 420
Db 362 VSLTRAQAMEKQDLAQLLEKDLQVQNEKINLRLMVLTSSTLQOELKAKRRRV 421
QY 421 TWCLGKINKMNSYADQFNPTNITTTKTHKLSINLAREIDESVCSSEDFSNLDTLSE 480
Db 422 TWCLGKINKMNSYADQFNPTNITTTKTHKLSINLAREIDESVCSSEDFSNLDTLSE 481
QY 481 IENWNPAT 487
Db 482 IENWNPAT 488

RESULT 2
US-09-150-867-1
; Sequence 1, Application US/09150867
; Patent No. 6645748
; GENERAL INFORMATION:
; APPLICANT: Wood, Kenneth W.
; APPLICANT: Sakowicz, Roman
; APPLICANT: Goldstein, Lawrence S.B.
; APPLICANT: Cleveland, Don W.
; TITLE OF INVENTION: Plus End-Directed Microtubule Motor Required for
; TITLE OF INVENTION: Chromosome Congression
; FILE REFERENCE: 18557C-0001100S
; CURRENT APPLICATION NUMBER: US/09/150,867
; CURRENT FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: US 60/058,645
; EARLIER FILING DATE: 1997-09-11
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2954
; TYPE: PRT
; ORGANISM: Xenopus sp.
; FEATURE:
; OTHER INFORMATION: Xenopus centromere-associated protein-E (XCENP-E)
; OTHER INFORMATION: member of the kinesin superfamily of microtubule
; OTHER INFORMATION: motor proteins
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (1)..(472)
; OTHER INFORMATION: kinesin like motor domain
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (473)..(2752)
; OTHER INFORMATION: rod domain
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (2753)..(2954)
; OTHER INFORMATION: tail domain
US-09-150-867-1

Query Match 63.8%; Score 1584; DB 4; Length 2954;
Best Local Similarity 64.0%; Pred. No. 4,7e-129; Indels 16; Gaps 6;
Matches 320; Conservative 68; Mismatches 96;

QY 1 AREGAVAVCVVRPLNSREESLGETAQVYWKTDNNVYQVDSKSFNDRVPHGNETTKN 60
Db 2 SEGDAVKVCVRPLNREQ--GDQANLQWKAGNNTISQVDGTSKSFNDRVFNESHSTQ 59
QY 61 VYEETAAPIIDSAIQYNGTIFAYGQTASGKYTMGSEDLHLGVIPRAHDFQIKKFP 120
Db 60 IYQELAVPIRSALQYNGTIFAYGQTSSGKYTMGTGPNLSLGIIPQAIQEVFKIIQETP 119
QY 121 DREFLLRVSYMEIYNETITDLCGQTKMKPLIREDVNRNVYVADLTERVYVTSEMA 180
Db 120 NREFLLRVSYMEIYNETVKDLCDRRKKPLREIREDFNRNVYVADLTERLVMPVPHIQW 179

```

```

QY 181 ITKGEKSRHYGETOMQRRSSRSHTIFRMILSREKGEPS---NCEGSVKVSHLNLVDLAG 237
Db 180 IKKGEKSRHYGETOMQRRSSRSHTIFRMIVESDRNDPTNSCNCDSGAVVSHLNLVDLAG 239
QY 238 SERAAQTGAAGVRLKEGCNINRSIFILGOVVKLSDGQGVGFNYRDSKLTRILQNSLGG 297
Db 240 SERAQTGAAGVRLKEGCNINRSIFILGOVVKLSDGQGVGFNYRDSKLTRILQNSLGG 299
QY 298 NPKTRIICITTPVSFDETLTALQFASAKYMKNTPPYNEVSTDEALKRYRKEIMDLKKQ 357
Db 300 NAKTVIICITTPVSFDETLTALQFASAKYMKNTPPYNEVSTDEALKRYRKEIMDLKKQ 359
QY 358 LE--EVSLETRAQAMEKQDLAQLLEKDLQVQNEKINLRLMVLTSSTLQOELKAK 415
Db 360 LENLESSETKAQAMAKEHTQLIAETIKQLHKEDEIRIWHLTNTVAVASSQES--QDQVRK 418
QY 416 RKRRVTWCLGKINKMNSYADQFN---IPTNITTTKTHKLSINLAREIDESVCSSEDF 471
Db 419 RKRRVTWAPGKIQNSLHASGVSDFDMLSRUPGNFKKAKFSDMPSFEIDDSVCTERSDF 478
QY 472 SNTLDTLS----EIEWNPAT 487
Db 479 DDALSMDNSNGIDAENLAS 498

RESULT 3
US-09-592-054-8
; Sequence 8, Application US/09592054
; Patent No. 6440684
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Finer, Jeffrey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Wood, Kenneth
; TITLE OF INVENTION: No. 6440684el motor proteins and methods for
; TITLE OF INVENTION: their use
; FILE REFERENCE: 1016
; CURRENT APPLICATION NUMBER: US/09/592,054
; CURRENT FILING DATE: 2000-07-20
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1234
; TYPE: PRT
; ORGANISM: Human
US-09-592-054-8

Query Match 28.4%; Score 705.5; DB 4; Length 1234;
Best Local Similarity 34.9%; Pred. No. 9.6e-53; Indels 65; Gaps 14;
Matches 182; Conservative 93; Mismatches 181;

QY 6 VAVCVVRPLNSREESLG-ETAQVYWKTDNNVYQVDSKSFNDRVPHGNETTKNVEE 64
Db 10 VRVALRCHPLVPKEISEGCMCLSFVPGETQVWGTD--KSTYDFVEDPCTEQEVFNK 67
QY 65 TAAPIDSAIQYNGTIFAYGQTASGKYTMG-----SEDLHLGVIPRAHDFQIK 117
Db 68 AVAPLIKGFYKYNATVLAAYGQTSGKYSMGCAVTAQENEPYGLIIPRVQLLFKEID 127
QY 118 KFPDREFLLRVSYMEIYNETITDLCGQTKMKPLIREDVNRNVYVADLTERVYVTSEMA 177
Db 128 QKSDSEFTLKVSYLEIYNEEILDLCLPSREKQAQINREDPKEGIKVLGTEKTLVALDT 187
QY 178 LKWIITKGEKSRHYGETKKNORSRSHTIFRMILSREKGEPSNCSGSKVSHLNLVDLAG 237
Db 188 VSCLEQGNNSRTVASTAMNSQSSRSHTAFTI---SLEQKKSKDKNSFR-SKLHLVDLAG 243
QY 238 SERAAQTGAAGVRLKEGCNINRSIFILGOVVKLSDGQGVGFNYRDSKLTRILQNSLGG 297
Db 244 SERQKTKTAGDRLEKGININRGLCLGNVISALGDDKAGSFVPIYRDSKLTRILQNSLGG 303
QY 298 NPKTRIICITTPV--SFDETLTALQFASAKYMKNTPPYNEVSTDEALKRYRKEIMDLK 355

```

```
Db 304 NSHTLMIACVSPADSNLEETLSTLRVADRARKKNKPIVN-----IDPHTAELNHLK 355
QY 356 KQLEEVSL-----ETRAQAMEKQDLAQLEEKD-----LQKVONEK 392
Db 356 QVQOQLOVLLQAHGGTLPGSINAPSENLSLMKKNQ--SLVBENEKLSRCLSKAAGQT 413
QY 393 IENLRLMTVSSSUTLQOELKAKRK--RRVTWCLGKLNK---MKNSNYADQFNIPNIT 446
Db 414 AQMLERIIITLQEQ---VNEKLNKLELRQHAACKLDLQKLVETLEQELKENVEILCNLQ 470
QY 447 TKTHKLSINLLRDRSVCSDESDFSNLTDLTSLSEIENWPAT 487
Db 471 QLITQLS-----DETVACTAAAIIDTAVBEEAQVETSPET 504

RESULT 4
US-09-595-684B-23
; Sequence 23, Application US/09595684B
; Patent No. 6544766
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Ohashi, Cara
; APPLICANT: Sakowicz, Roman
; APPLICANT: Vaisberg, Eugeni
; APPLICANT: Wood, Kenneth
; APPLICANT: Yu, Ming
; TITLE OF INVENTION: Human kinesins and methods of producing
; and purifying human kinesins
; FILE REFERENCE: cytop036
; CURRENT APPLICATION NUMBER: US/09/595,684B
; CURRENT FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: 09/295,612
; PRIOR FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 1231
; TYPE: PRT
; ORGANISM: Human
US-09-595-684B-23

Query Match 28.4%; Score 704; DB 4; Length 1231;
Best Local Similarity 35.4%; Pred. No. 1.3e-52;
Matches 183; Conservative 85; Mismatches 191; Indels 58; Gaps 12;

QY 6 VAVCVVRPLNSRRESLG-ETAAQVWKTNNVIYQVDSKSNFDRVFHGNETTNNVYEE 64
Db 10 VRVALRCRLVPKEISEGCMCLSFVPGEPQVWVGTD--KSFTYDFVDPSTEQEVFNT 67
QY 65 IAAPIIDSAIQYNGTIFAYGOTASGKTYTMG-----SDHLGVIPRAIHDIFQKIK 117
Db 68 AVAPLKGFKYNATVLAQGTGSGKTYSMGAVTAEOENPTVGVIPRVQLLFKEID 127
QY 118 KPPDFEFLRVSMEIYNETITDLGCTQKMKPLIREDVNRNRYVADLTREVVTSEMA 177
Db 128 KKSDFEFLKVSLEYNEEILLCLPCSKREKQINREDPKGKIIVGTEKTVLVALDT 187
QY 178 LKWTIKGKSRHYGKTKNORSRSHITFRMLESREKGEPSNCRGSKVSHNLVLDAG 237
Db 188 VSCLEQGNNSRTVASTAMNSQSRSRSHAFIT---SLEQKKSDKNSSFR-SKLHLVDLAG 243
QY 238 SERAAQTGAAGVRLKEGNCINRSLFILGOVKKLSDGQGVGFNRYRDSKLTILQNSLGG 297
Db 244 SERQKKTVAEGDLKEGININRGLCLGNVISALGDDKGGGPAPYRDSKLTLLQDSLGG 303
QY 298 NPKTRIICTITPV--SFDETTLTALQASTAKYMKNTPYVNEVSTDEALLKRYKEIMDLK 355
Db 304 NSHTLMIACVSPADSNLEETLSTLRVADRARKKNKPIVN-----IDPHTAELNHLK 355
QY 356 KQLEEVSL-----ETRAQAMEKQDLAQLEEKD-----LQKVONEK 392
Db 356 QVQOQLOVLLQAHGGTLPGSINAPSENLSLMKKNQ--SLVBENEKLSRCLSKAAGQT 413
```

```
QY 397 TRML-----VTSSSUTLQOELKAKKRRTVWC---LGKINKMKNSNYADQFNIPNITTKTH 450
Db 414 AQMLERIIITQOANERKNNAKLELRQHAACKLDLQKLVETLDQELKENVEILCNLQKIT 473
QY 451 KLSINLLRDRSVCSDESDFSNLTDLTSLSEIENWPAT 487
Db 474 QLS-----DETVACTAAAIIDTAVBEEAQVETSPET 503
```

RESULT 5

```
US-09-541-782-8
; Sequence 8, Application US/09541782
; Patent No. 6284480
; GENERAL INFORMATION:
; APPLICANT: Nislow, Corey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Beraud, Christophe
; TITLE OF INVENTION: Antifungal Assay
; FILE REFERENCE: 1015
; CURRENT APPLICATION NUMBER: US/09/541,782
; CURRENT FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1066
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-541-782-8
```

Query Match 28.3%; Score 703; DB 3; Length 1066;
Best Local Similarity 34.4%; Pred. No. 1.3e-52;
Matches 187; Conservative 93; Mismatches 183; Indels 80; Gaps 17;

```
QY 6 VAVCVVRPLNSRRESL--GETAQVWKTNNVIYQVDS--SKSNFDRVHGNETTNNV 61
Db 20 IQVYVRPLNSRERCIRSAEVVDVVGPEVTRHLDLKLTKKFTFDRSGFSPESQCDV 79
QY 62 YEEIAPIIDSAIQYNGTIFAYGOTASGKTYTMGSE-----DHLGVIPRAIH 110
Db 80 YSVVSPLEIEVLNGYNGCTVFAYGQTGKTHTWVGNETAELKSSWEDSDIGIIPRLS 139
QY 111 DIFQIKKFPDFEFLRVSMEIYNETITDLGCTQKMKPLIREDVNR--NVYVADLTTEE 169
Db 140 HUFDELRVM--EVEYTMRIYSLEYNEELCDLSTDDTKIRIFDDSTKKGSVIIQGLEEI 198
QY 170 VVYTSMAKWTIKGKSRHYGETKNORSRSHITFRMLESREKGEPSNCRGSKVYSH 229
Db 199 PVHSKDDVYKLEKGERKRTATILMNAQSSRSHVFSIVVHIRENGIEG--EDMKIKGK 256
QY 230 LNLVDLAGSRAAQTG--AAGVRLKEGNCINRSLFILGOVKKLSDGQGVGFNRYRDSKIT 288
Db 257 LNLVDLAGSENVSAGNEKGRVRETVNINQSLLTGRVITLVDR--APHVPYRESKIT 314
QY 289 RILQNSLGCNPKTRIICTITP--VSFDETLTALQASTAKYMKNTPYVNEVSTDEALLKR 346
Db 315 RLLQESLGGRTKTSIIATISPGKDIETSLTLEYAHRKNIQNKPEVNOQLTKKTVLKE 374
QY 347 YRKEIMDLKKQL-----EVSLETRAQAMEKQDLAQLEB--KDLLQKVQ 389
Db 375 YTEIDKLRDLMAARDKNGIYLAEEYGETLKLSEQNRELNKMLLKALKDLQNK 434
QY 390 N-----EKIENLTR---MLVSSSUTLQOELKAKKRRTVWCILGKINKMKNSNY 435
Db 435 KIFSEVSNLSVEKTOELKKTEENLLNTKGTLLTKKVLTKKRY-----KEKKELV 486
QY 436 ADQFNIPNITTK-----THKU--SINLREIDESVCSDESDFSNLT--DTL 478
Db 487 ASHMKTEQVLTQAOEILAAADLATDTHQLHGTIERREDELDEKIRRSQDFKORMQDL 546
QY 479 SEI 481
Db 547 EMI 549
```

GENERAL INFORMATION:
; APPLICANT: Nislow, Corey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Beraud, Christophe
; TITLE OF INVENTION: Antifungal Assay
; FILE REFERENCE: 1015
; CURRENT APPLICATION NUMBER: US/10/270,085
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: US/09/723,820
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 09/541,782
; PRIOR FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1066
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-270-085-8

Query Match 28.3%; Score 703; DB 4; Length 1066;
Best Local Similarity 34.4%; Pred. No. 1.3e-52;
Matches 187; Conservative 93; Mismatches 183; Indels 80; Gaps 17;

Qy 6 VAVCVVRPLNSREESL--GETAQVYWKTDNNVIYQVDG--SKSFNDRVPHGNETTKNV 61
Db 20 IQVTVRPLNSRERCIRSAEVDVVGPREVTRHTLDSKLTKKFTFDRSGPESKQCDV 79
Qy 62 YEEIAAPLIDSAIOGYNGTIFAYGQTASGKTYTMGSE-----DHLGVIPIRAIH 110
Db 80 YSVVVSPLIEVLNGYNTVFAYGQTGKTHTWVGNETAELKSSWEDSDIGIIPRALS 139
Qy 111 DIFOKIKKFPDRFLLRVSMYIYNETITDLCGTQKMKPLIREDVNR--NVYVADLTEE 169
Db 140 HLFDELMM--EVEYTMRIISYLELYNEELCDLLSDTDTTKIRIFDDSTKKGSVIIQGLEEI 198
Qy 170 VVYTSMALKWIITGKSRHYGETKMNORSRSHITIFRMILESEKGPSPNCEGSKVYSH 229
Db 199 PVHSKDDVYKLEKGERKRTATTLMAQSSRSHTVFSIVVHIRENGIEG--EDMLKIGK 256
Qy 230 LNLVDLAGSERRAAQTG--AAGVRLKEGNCNINRSLFILQVIVKLSGQVGGFINVRDSKLT 288
Db 257 LNLVDLAGSENVSAGNEKGIRVRETVMNQSLTLGRVITVDR--APHVPYRESKLT 314
Qy 289 RILQNSLGNPKTRIICTITP--VSFDETLTALQFASTAKYMKNTPYVNEVSTDEALLKR 346
Db 315 RLQESLGRTKTSIIATISPGHKDIBETLSTLEYAHRAKNIQNKPEVNQKLTKKTVLKE 374
Qy 347 YRKEIMDLKKQI-----EEVSLETRAQAMEKQDLAQLEE--KDLQKVQ 389
Db 375 YTEIDLKRLDMAARDKNGIYLAETVGEITLKLESQNRRELNEKMLLKALKDBLQNK 434
Qy 390 N-----EKIENLTR-----MLVTSSTLQOELKAKRRRVTCWLGKINKMKNY 435
Db 435 KIPSEVSMLEKTOELKKEKTEENLNTKGTLLTKVLTKKRY-----KEKKELV 486
Qy 436 ADQFNIPNTITK-----THKL---SINLREIDESVCSSESVFNTL--DTL 478
Db 487 ASHMKTEQVLTQAOEILAAADLATDDTHLGHGTIERREBELDEKIRRSQDFKDRMQDNL 546
Qy 479 SEI 481
Db 547 EMI 549

RESULT 8
US-09-592-054-4
; Sequence 4, Application US/09592054
; Patent No. 6440684
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Finer, Jeffrey
; APPLICANT: Sakowicz, Roman

US-09-723-820-8
; Sequence 8, Application US/09723820
; Patent No. 6468760
; GENERAL INFORMATION:
; APPLICANT: Nislow, Corey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Beraud, Christophe
; TITLE OF INVENTION: Antifungal Assay
; FILE REFERENCE: 1015
; CURRENT APPLICATION NUMBER: US/09/723,820
; CURRENT FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 09/541,782
; PRIOR FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1066
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-723-820-8

Query Match 28.3%; Score 703; DB 4; Length 1066;
Best Local Similarity 34.4%; Pred. No. 1.3e-52;
Matches 187; Conservative 93; Mismatches 183; Indels 80; Gaps 17;

Qy 6 VAVCVVRPLNSREESL--GETAQVYWKTDNNVIYQVDG--SKSFNDRVPHGNETTKNV 61
Db 20 IQVTVRPLNSRERCIRSAEVDVVGPREVTRHTLDSKLTKKFTFDRSGPESKQCDV 79
Qy 62 YEEIAAPLIDSAIOGYNGTIFAYGQTASGKTYTMGSE-----DHLGVIPIRAIH 110
Db 80 YSVVVSPLIEVLNGYNTVFAYGQTGKTHTWVGNETAELKSSWEDSDIGIIPRALS 139
Qy 111 DIFOKIKKFPDRFLLRVSMYIYNETITDLCGTQKMKPLIREDVNR--NVYVADLTEE 169
Db 140 HLFDELMM--EVEYTMRIISYLELYNEELCDLLSDTDTTKIRIFDDSTKKGSVIIQGLEEI 198
Qy 170 VVYTSMALKWIITGKSRHYGETKMNORSRSHITIFRMILESEKGPSPNCEGSKVYSH 229
Db 199 PVHSKDDVYKLEKGERKRTATTLMAQSSRSHTVFSIVVHIRENGIEG--EDMLKIGK 256
Qy 230 LNLVDLAGSERRAAQTG--AAGVRLKEGNCNINRSLFILQVIVKLSGQVGGFINVRDSKLT 288
Db 257 LNLVDLAGSENVSAGNEKGIRVRETVMNQSLTLGRVITVDR--APHVPYRESKLT 314
Qy 289 RILQNSLGNPKTRIICTITP--VSFDETLTALQFASTAKYMKNTPYVNEVSTDEALLKR 346
Db 315 RLQESLGRTKTSIIATISPGHKDIBETLSTLEYAHRAKNIQNKPEVNQKLTKKTVLKE 374
Qy 347 YRKEIMDLKKQI-----EEVSLETRAQAMEKQDLAQLEE--KDLQKVQ 389
Db 375 YTEIDLKRLDMAARDKNGIYLAETVGEITLKLESQNRRELNEKMLLKALKDBLQNK 434
Qy 390 N-----EKIENLTR-----MLVTSSTLQOELKAKRRRVTCWLGKINKMKNY 435
Db 435 KIPSEVSMLEKTOELKKEKTEENLNTKGTLLTKVLTKKRY-----KEKKELV 486
Qy 436 ADQFNIPNTITK-----THKL---SINLREIDESVCSSESVFNTL--DTL 478
Db 487 ASHMKTEQVLTQAOEILAAADLATDDTHLGHGTIERREBELDEKIRRSQDFKDRMQDNL 546
Qy 479 SEI 481
Db 547 EMI 549

RESULT 7
US-10-270-085-8
; Sequence 8, Application US/10270085
; Patent No. 6627408

APPLICANT: Wood, Kenneth
TITLE OF INVENTION: No. 6440684el motor proteins and methods for
FILE REFERENCE: 1016
CURRENT APPLICATION NUMBER: US/09/592,054
CURRENT FILING DATE: 2000-07-20
NUMBER OF SEQ ID NOS: 8
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4
LENGTH: 522
TYPE: PRT
ORGANISM: Human
US-09-592-054-4

Query Match 28.3%; Score 702; DB 4; Length 522;
Best Local Similarity 35.8%; Pred. No. 5.4e-53;
Matches 181; Conservative 83; Mismatches 187; Indels 54; Gaps 12;
QY 6 VAVCVVRPLNSREESLG-ETAQVYWKTDNNVIYQVDSKSFNFDRVPHGNETTKNVYEE 64
DB 30 VRVALRCRPLVPKEISEGCMCLSFVPGEPVQVVGTD--KSTYDFVDFPSTEQEEVNT 87
QY 65 IAAPIDSAIOGYNGTIFAYGOTASGKTVMWG-----SEDLGVIPRAIHDIQKIK 117
DB 88 AVAPLKGKGVKGNATVLAAYGOTGSKTYSMGAYTAEQENPTVGVIPRVQLLFKEID 147
QY 118 KPDPREFLRVSMETYNETITDLCGTOKMKPLIREDVNRNYYVADLTTEEVTSEMA 177
DB 148 KKSDFEFTLVKSYLEIYNEEILDLCPSREKAIQINREDPKGKIVGLTEKTVLVALDT 207
QY 178 LKMITKGEKSRHYGTOMNORSRSHITPRMILESEKGEPCNCGSVKVSHLNLDVLAG 237
DB 208 VSCLEQGNNSRTVASTAMNSQSSRSHAITISLEQKKSD----KNSFSRKLHLVDLAG 263
QY 238 SRAAQTGAAGVRLKEGNCNINSLFILGOVKKLSGQVGGFINVRDSKLTILQNSLGG 297
DB 264 SERQKTKAEGDRLKEGININRGLLGNVISALGDDKGGFVPYRDSKLTLLQDSLGG 323
QY 298 NPKTRIICTITPV--SFDETLALOFASTAKYMKNTPYVNEVSTDEALLKRYRKEIMDLK 355
DB 324 NSHTLMIACVSPADSNLEETLNTLAYADRARKIKKPIVN-----IDPQTAEINHLK 375
QY 356 KQLEEVSL-----ETRAQMEKQDLAQLLEKDLQKVNKEIENL 396
DB 376 QVQOQLVLLQAHGGLPGSTVTPSENLQSLMEKNQ--SLVEENEKLSRGLSEAAGT 433
QY 397 TRLM-----VTSSSLTIQOELKAKRKRVTWCLGKINK----MKNVYADQFNPTWTTK 448
DB 434 AQMLERIIWTEQANEMKNAKLEELRQHAACKLDLQKLVTLEDBELKENVEIICNLQQLI 493
QY 449 THKLSINLLREIDESVCSSESDVFSN 473
DB 494 TQKLA--ALEGTEQKLISEEDLIEH 517

RESULT 9
US-09-592-054-2
Sequence 2, Application US/09592054
Patent No. 6440684
GENERAL INFORMATION:
APPLICANT: Beraud, Christophe
APPLICANT: Finer, Jeffrey
APPLICANT: Sakowicz, Roman
APPLICANT: Wood, Kenneth
TITLE OF INVENTION: No. 6440684el motor proteins and methods for
FILE REFERENCE: 1016
CURRENT APPLICATION NUMBER: US/09/592,054
CURRENT FILING DATE: 2000-07-20
NUMBER OF SEQ ID NOS: 8
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 1232

TYPE: PRT
ORGANISM: Human
US-09-592-054-2

Query Match 28.1%; Score 696.5; DB 4; Length 1232;
Best Local Similarity 34.9%; Pred. No. 5.9e-52;
Matches 181; Conservative 86; Mismatches 192; Indels 59; Gaps 12;
QY 6 VAVCVVRPLNSREESLG-ETAQVYWKTDNNVIYQVDSKSFNFDRVPHGNETTKNVYEE 64
DB 10 VRVALRCRPLVPKEISEGCMCLSFVPGEPVQVVGTD--KSTYDFVDFPSTEQEEVNT 67
QY 65 IAAPIDSAIOGYNGTIFAYGOTASGKTVMWG-----SEDLGVIPRAIHDIQKIK 117
DB 68 AVAPLKGKGVKGNATVLAAYGOTGSKTYSMGAYTAEQENPTVGVIPRVQLLFKEID 127
QY 118 KPDPREFLRVSMETYNETITDLCGTOKMKPLIREDVNRNYYVADLTTEEVTSEMA 177
DB 128 KKSDFEFTLVKSYLEIYNEEILDLCPSREKAIQINREDPKGKIVGLTEKTVLVALDT 187
QY 178 LKMITKGEKSRHYGTOMNORSRSHITPRMILESEKGEPCNCGSVKVSHLNLDVLAG 237
DB 188 VSCLEQGNNSRTVASTAMNSQSSRSHAITISLEQKKSDKNSFSRKLHLVDLAG 243
QY 238 SRAAQTGAAGVRLKEGNCNINSLFILGOVKKLSGQVGGFINVRDSKLTILQNSLGG 297
DB 244 SERQKTKAEGDRLKEGININRGLLGNVISALGDDKGGFVPYRDSKLTLLQDSLGG 303
QY 298 NPKTRIICTITPV--SFDETLALOFASTAKYMKNTPYVNEVSTDEALLKRYRKEIMDLK 355
DB 304 NSHTLMIACVSPADSNLEETLNTLAYADRARKIKKPIVN-----IDPQTAEINHLK 355
QY 356 KQLEEVSL-----ETRAQMEKQDLAQLLEKDLQKVNKEIENL 396
DB 356 QVQOQLVLLQAHGGLPGSTVTPSENLQSLMEKNQ--SLVEENEKLSRGLSEAAGT 413
QY 397 TRLM-----VTSSSLTIQOELKAKRKRVTWCLGKINK----MKNVYADQFNPTWTTK 449
DB 414 AQMLERIIWTEQANEMKNAKLEELRQHAACKLDLQKLVTLEDBELKENVEIICNLQQLI 473
QY 450 HKLSINLLREIDESVCSSESDVFSNLTDLSEIEMNPAT 487
DB 474 TQLS-----DETVMCAAALDAITVEQEAQVETSPET 504

RESULT 10
US-09-592-054-6
Sequence 6, Application US/09592054
Patent No. 6440684
GENERAL INFORMATION:
APPLICANT: Beraud, Christophe
APPLICANT: Finer, Jeffrey
APPLICANT: Sakowicz, Roman
APPLICANT: Wood, Kenneth
TITLE OF INVENTION: No. 6440684el motor proteins and methods for
FILE REFERENCE: 1016
CURRENT APPLICATION NUMBER: US/09/592,054
CURRENT FILING DATE: 2000-07-20
NUMBER OF SEQ ID NOS: 8
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6
LENGTH: 473
TYPE: PRT
ORGANISM: Human
US-09-592-054-6

Query Match 28.0%; Score 695; DB 4; Length 473;
Best Local Similarity 39.7%; Pred. No. 1.9e-52;
Matches 172; Conservative 70; Mismatches 155; Indels 36; Gaps 10;
QY 6 VAVCVVRPLNSREESLG-ETAQVYWKTDNNVIYQVDSKSFNFDRVPHGNETTKNVYEE 64


```

Db 356 KEKNARLKGKVEKLEIELARWRAGETVKAEEQINMEDLMSEAGTPNLEVE-AAQTAAAEAA 414
QY 406 LTLOQELKA-----KXKRVVWCLGKINKM-----KNSNYADQF-----N 440
Db 415 LAAQRTALANASAVVNEQALATECERLYOQLODKDEBEINQOQYAEQLEQVMEQBE 474
QY 441 IPTNITTKHLSINLLREIDSVCSSESVFNTSLDITLSEIENW 484
Db 475 LIANARREYETLOSEWARIQENESAKEEV-KEVLQALEELAVN 517

RESULT 13
US-09-572-191-2
; Sequence 2, Application US/09572191
; Patent No. 6355466
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Sakowicz, Roman
; APPLICANT: Wood, Kenneth
; TITLE OF INVENTION: No. 6355466el motor proteins and methods for
; TITLE OF INVENTION: their use
; FILE REFERENCE: 1017
; CURRENT APPLICATION NUMBER: US/09/572,191
; PRIOR FILING DATE: 2000-05-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1388
; TYPE: PRT
; ORGANISM: Human
US-09-572-191-2
Query Match 26.4%; Score 656; DB 4; Length 1388;
Best Local Similarity 37.8%; Pred. No. 2.4e-48;
Matches 201; Conservative 69; Mismatches 168; Indels 94; Gaps 20;
QY 2 BEGAVAVCVRVRLNRSREESL-GETAQVYWKTDNNVIYQVDGS-----KSFNFD 49
Db 23 EGDALKVFRIRPPAERSGSADGE-----QNLCLSVLSSTSLRLHSNPEPTFTFD 73
QY 50 RVFHGNETTKNVEIEAIPIDSAIOGNGTIFAYGQTASGKTYTMG-----SEHIL 102
Db 74 HVADVDTTQESVFATVAKSVESCMGNGTIFAYGQTSGKTFMTMGSPESDNFSNLR 133
QY 103 GVIPRAIHDF-----QKIKFPDREFLLRVSYMEIYNETITDLCGQTKMKPLIREDV 157
Db 134 GVIPRSFYLFLDREKEKAGAGKSFLEKCSFIEYNEIYDILL--DSASAGLYLREHI 191
QY 158 NRVVYVADLTBEVVYVTSMAKWKITGKSRHYGETKMNQSRSSHTIFRMIRESKGE 217
Db 192 KKGVFVVGAVQVTSAAEAQVLSGGWRNRVASTSMNRESSRSHAVFTTIESMEK-- 249
QY 218 PSNCEGSVKVSHNLVDLAGSRAAQTGAAGVRLKEGNNINRSLFILQGVIKKLSGQVG 277
Db 250 -SNEIVNRTSLNLLVDLAGSERQDTHAEGMLKEAGNINRSLSCGQVITALVD--VG 306
QY 278 G-----FINYRDSKLTILQNSLGGNPKTRIICITTPVS--FDETLTALQFASTAKYMKNT 331
Db 307 NGKQHVYCYRDSKLTFLRLDSLGGNAKTAIIANVHPGSRGCFGETLSTLNFAQAKLIKNK 366
QY 332 PYVNE-----VSTDEALKRYRKEIMDLKK-QLEEVSLETR-----AQAMEKDQLAQLL-- 379
Db 367 AVNEDTQGNVSQLQAEVVKRLEQALASGQTPPEPFLTRDKKNTNMYEYFQEAFLFFK 426
QY 380 ----BEKOLLQVQNEKIENLRMLVTSSTLTLOQELKAKKRRVTCWLGKINKMKNSNY 435
Db 427 KSEQEKSLIEKV-----TQLEDLTLLKKEKFI-----QSNKMIVKFR 463
QY 436 ADQFNIPNITTKHLSI-NLLREIDSVCSSESVFNTSLDITLSE-LEWNP 485
Db 464 EDQI-----IRLEKHLKESRGGLPEEQDRLISE---LRNEIQTREIEHP 508

```

```

Query Match 26.4%; Score 656; DB 4; Length 1388;
Best Local Similarity 37.8%; Pred. No. 2.4e-48;
Matches 201; Conservative 69; Mismatches 168; Indels 94; Gaps 20;
QY 2 BEGAVAVCVRVRLNRSREESL-GETAQVYWKTDNNVIYQVDGS-----KSFNFD 49
Db 23 EGDALKVFRIRPPAERSGSADGE-----QNLCLSVLSSTSLRLHSNPEPTFTFD 73
QY 50 RVFHGNETTKNVEIEAIPIDSAIOGNGTIFAYGQTASGKTYTMG-----SEHIL 102
Db 74 HVADVDTTQESVFATVAKSVESCMGNGTIFAYGQTSGKTFMTMGSPESDNFSNLR 133
QY 103 GVIPRAIHDF-----QKIKFPDREFLLRVSYMEIYNETITDLCGQTKMKPLIREDV 157
Db 134 GVIPRSFYLFLDREKEKAGAGKSFLEKCSFIEYNEIYDILL--DSASAGLYLREHI 191
QY 158 NRVVYVADLTBEVVYVTSMAKWKITGKSRHYGETKMNQSRSSHTIFRMIRESKGE 217
Db 192 KKGVFVVGAVQVTSAAEAQVLSGGWRNRVASTSMNRESSRSHAVFTTIESMEK-- 249
QY 218 PSNCEGSVKVSHNLVDLAGSRAAQTGAAGVRLKEGNNINRSLFILQGVIKKLSGQVG 277
Db 250 -SNEIVNRTSLNLLVDLAGSERQDTHAEGMLKEAGNINRSLSCGQVITALVD--VG 306
QY 278 G-----FINYRDSKLTILQNSLGGNPKTRIICITTPVS--FDETLTALQFASTAKYMKNT 331
Db 307 NGKQHVYCYRDSKLTFLRLDSLGGNAKTAIIANVHPGSRGCFGETLSTLNFAQAKLIKNK 366
QY 332 PYVNE-----VSTDEALKRYRKEIMDLKK-QLEEVSLETR-----AQAMEKDQLAQLL-- 379
Db 367 AVNEDTQGNVSQLQAEVVKRLEQALASGQTPPEPFLTRDKKNTNMYEYFQEAFLFFK 426
QY 380 ----BEKOLLQVQNEKIENLRMLVTSSTLTLOQELKAKKRRVTCWLGKINKMKNSNY 435
Db 427 KSEQEKSLIEKV-----TQLEDLTLLKKEKFI-----QSNKMIVKFR 463
QY 436 ADQFNIPNITTKHLSI-NLLREIDSVCSSESVFNTSLDITLSE-LEWNP 485
Db 464 EDQI-----IRLEKHLKESRGGLPEEQDRLISE---LRNEIQTREIEHP 508

```

```

RESULT 14
US-09-723-262-2
; Sequence 2, Application US/09723262
; Patent No. 6379912
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Sakowicz, Roman
; APPLICANT: Wood, Kenneth
; TITLE OF INVENTION: No. 6379912el motor proteins and methods for
; TITLE OF INVENTION: their use
; FILE REFERENCE: 1017
; CURRENT APPLICATION NUMBER: US/09/723,262
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: US 09/572,191
; PRIOR FILING DATE: 2000-05-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1388
; TYPE: PRT
; ORGANISM: Human
US-09-723-262-2
Query Match 26.4%; Score 656; DB 4; Length 1388;
Best Local Similarity 37.8%; Pred. No. 2.4e-48;
Matches 201; Conservative 69; Mismatches 168; Indels 94; Gaps 20;
QY 2 BEGAVAVCVRVRLNRSREESL-GETAQVYWKTDNNVIYQVDGS-----KSFNFD 49
Db 23 EGDALKVFRIRPPAERSGSADGE-----QNLCLSVLSSTSLRLHSNPEPTFTFD 73
QY 50 RVFHGNETTKNVEIEAIPIDSAIOGNGTIFAYGQTASGKTYTMG-----SEHIL 102
Db 74 HVADVDTTQESVFATVAKSVESCMGNGTIFAYGQTSGKTFMTMGSPESDNFSNLR 133
QY 103 GVIPRAIHDF-----QKIKFPDREFLLRVSYMEIYNETITDLCGQTKMKPLIREDV 157
Db 134 GVIPRSFYLFLDREKEKAGAGKSFLEKCSFIEYNEIYDILL--DSASAGLYLREHI 191
QY 158 NRVVYVADLTBEVVYVTSMAKWKITGKSRHYGETKMNQSRSSHTIFRMIRESKGE 217
Db 192 KKGVFVVGAVQVTSAAEAQVLSGGWRNRVASTSMNRESSRSHAVFTTIESMEK-- 249
QY 218 PSNCEGSVKVSHNLVDLAGSRAAQTGAAGVRLKEGNNINRSLFILQGVIKKLSGQVG 277
Db 250 -SNEIVNRTSLNLLVDLAGSERQDTHAEGMLKEAGNINRSLSCGQVITALVD--VG 306
QY 278 G-----FINYRDSKLTILQNSLGGNPKTRIICITTPVS--FDETLTALQFASTAKYMKNT 331
Db 307 NGKQHVYCYRDSKLTFLRLDSLGGNAKTAIIANVHPGSRGCFGETLSTLNFAQAKLIKNK 366
QY 332 PYVNE-----VSTDEALKRYRKEIMDLKK-QLEEVSLETR-----AQAMEKDQLAQLL-- 379
Db 367 AVNEDTQGNVSQLQAEVVKRLEQALASGQTPPEPFLTRDKKNTNMYEYFQEAFLFFK 426
QY 380 ----BEKOLLQVQNEKIENLRMLVTSSTLTLOQELKAKKRRVTCWLGKINKMKNSNY 435
Db 427 KSEQEKSLIEKV-----TQLEDLTLLKKEKFI-----QSNKMIVKFR 463
QY 436 ADQFNIPNITTKHLSI-NLLREIDSVCSSESVFNTSLDITLSE-LEWNP 485
Db 464 EDQI-----IRLEKHLKESRGGLPEEQDRLISE---LRNEIQTREIEHP 508

```

```

RESULT 15
US-09-723-219-2
; Sequence 2, Application US/09723219
; Patent No. 6391613
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Sakowicz, Roman
; APPLICANT: Wood, Kenneth
; TITLE OF INVENTION: No. 6391613el motor proteins and methods for

```

```

; TITLE OF INVENTION: their use
; FILE REFERENCE: 1017
; CURRENT APPLICATION NUMBER: US/09/723,219
; CURRENT FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: US 09/572,191
; PRIOR FILING DATE: 2000-05-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1388
; TYPE: PRT
; ORGANISM: Human
; US-09-723-219-2

Query Match          26.4%; Score 656; DB 4; Length 1388;
Best Local Similarity 37.8%; Pred. No. 2.4e-48;
Matches 201; Conservative 69; Mismatches 168; Indels 94; Gaps 20;

QY 2 EGAVAVCVVRVRLNREESL-GETAQVYWKTDNNVIYQVDS-----KGFNFD 49
Db 23 EGDAIKVFVRAIRPPAERSGSADGE-----QNLCLSVLSSTSLRLHSNPEPKFTFD 73
QY 50 RVFHGNETTKNVVEETAAPIDSALQCYNGTIFAYGOTASGKTYTMWG-----SEDHL 102
Db 74 HVADVDVTTQESVFATVAKSIVFSCMGSYNGTIFAYGOTGSGKTFMMGPSESNFSLNR 133
QY 103 GVIPRAIHDIF-----QKIKFPDPREFLRVSYMEIYNETITDLCGTQKMKPLIREDV 157
Db 134 GVIPRSEYLFSLIDREKEKAGAKSFLCKSFIEIYNEQIYDLL--DSASAGLYLREHI 191
QY 158 NRNVYVADLTVEVVYTSEMALKWITKGEGRHYGETKQNRQSRSHYTFRMILESRKGE 217
Db 192 KGVFVVGAVEQVVTSAEAYQVLSGGWRNRVASTSMNRESSRSHAVFTTIESMEK-- 249
QY 218 PSNCEGSKVYSHLNLVDLAGSRAAQTGAAGVRLKEGCNINRSLFILGOVIKKLSGQVG 277
Db 250 -SNEIVNIRTSLLNLVDLAGSERQKDTAEGMRLEKAGNINRSLSCIGQVITALVD--VG 306
QY 278 G----FINYRDSKLTILQNSLGNPKTRIITTPVS--FDETTLALQFPASTAKYMKNT 331
Db 307 NGQQRHVCYRDSKLTFLRDSLGNKATAIANVHPSRCFGETLSTLNFQAQAKLIKX 366
QY 332 PYVNE-----VSTDEALLKRYKEIMDLKX-QLEEVSLETR----AOMEXDQLAQL-- 379
Db 367 AVNEDTQGNVSQLQAEVKRLKEQLAELASQTPPESFLTRDKKKTNYMEYFOEAMLFFK 426
QY 380 ----EKKDLQCKVQNEKIENLTRMLVTSSSLTLQOELKAKRGRVTCWCLGKINKMKNSY 435
Db 427 KSEQEKSLIEK-----TQLEDLTLKKEFI-----QSNRMIVKFR 463
QY 436 ADQFNIPNTITTKTKLSI-NLLREIDESVCSDFSNFTIDTISE-IEWNP 485
Db 464 EDQI-----IRLEKLHESRGGLFEEQDRLLSE---LRNEIQTLREQIEHHP 508

```

Search completed: July 29, 2004, 09:42:49
Job time : 10.6277 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 09:39:47 ; Search time 24.324 Seconds
(without alignments)
6280.361 Million cell updates/sec

Title: US-10-045-631b-88_COPY_2_488

Perfect score: 2483

Sequence: 1 AEEGAVAVCVVRPLNSREE.....SDVFNLTDLTSLIEWNPAT 487

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/2/pubpaa/US05_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/2/pubpaa/US09C_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/2/pubpaa/US10C_NEW_PUB.pep:*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	973.5	39.2	1382	16	US-10-437-963-176714
2	878	35.4	694	12	Sequence 176714, A
3	750.5	30.2	1006	16	Sequence 59725, A
4	718.5	28.9	420	16	Sequence 147933, A
5	717.5	28.9	1237	15	Sequence 114373, A
6	714.5	28.8	955	12	Sequence 33, Appl
7	713.5	28.7	1232	14	Sequence 62672, A
8	713.5	28.7	1232	16	Sequence 670, App
9	713.5	28.7	1235	15	Sequence 2153, App
10	710.5	28.6	956	16	Sequence 8, Appl
11	710.5	28.6	1029	16	Sequence 182113, A
12	702.5	28.3	1232	14	Sequence 2, Appl
13	702.5	28.3	1232	14	Sequence 664, App
14	691.5	27.8	965	16	Sequence 689, App
15	689	27.7	672	16	Sequence 169150, A
					Sequence 1664, App

16	671.5	27.0	966	12	US-10-425-114-62697
17	664.5	26.8	1031	14	Sequence 62697, A
18	664.5	26.8	1031	15	Sequence 24, Appl
19	663.5	26.7	406	15	Sequence 113, App
20	660.5	26.6	517	15	Sequence 3789, App
21	660.5	26.6	975	14	Sequence 12881, A
22	660.5	26.6	975	15	Sequence 19, Appl
23	658	26.5	757	15	Sequence 108, App
24	658	26.5	757	12	Sequence 366, App
25	656	26.4	1388	14	Sequence 316, App
26	656	26.4	1388	12	Sequence 2, Appl
27	656	26.4	1388	15	Sequence 82, Appl
28	656	26.4	1388	16	Sequence 164, App
29	654.5	26.4	963	14	Sequence 20, Appl
30	654.5	26.4	963	15	Sequence 109, App
31	653.5	26.3	963	14	Sequence 22, Appl
32	653.5	26.3	963	15	Sequence 111, App
33	652.5	26.3	821	16	Sequence 114370, A
34	650	26.2	1362	12	Sequence 2, Appl
35	650	26.2	1826	16	Sequence 1477, App
36	648.5	26.1	1022	16	Sequence 133106, A
37	643.5	25.9	2552	16	Sequence 129734, A
38	642.5	25.9	967	14	Sequence 21, Appl
39	642.5	25.9	967	15	Sequence 110, App
40	640	25.8	957	14	Sequence 16, Appl
41	640	25.8	957	15	Sequence 105, App
42	638.5	25.7	1011	12	Sequence 24, Appl
43	638.5	25.7	1011	12	Sequence 58, Appl
44	638.5	25.7	1032	14	Sequence 26, Appl
45	638.5	25.7	1032	15	Sequence 115, App

ALIGNMENTS

RESULT 1

US-10-437-963-176714
; Sequence 176714, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 176714
; LENGTH: 1382
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_74437C.1.pep
US-10-437-963-176714

Query Match	39.2%	Score	973.5	DB	16	Length	1382
Best Local Similarity	46.7%	Pred. No.	4e-74				
Matches	232	Conservative	74	Mismatches	140	Indels	51
Gaps	15						
QY	6	VAVCVVRPLNSRESLCEATQVYWKTDNNVI-YQVDGSKGFNDRVFHGNKTVYEE	64				
Db	4	IHVAVRPLTA--EDAGSSP---WRVSGNAIALSTQPSIRFEDRIEGECRTADVGA	58				
QY	65	IAAPIIDSAIGVNGTIPAYGQTASGKTYTMWGSEDLHGVIPTRAIHDIQKIKFPDREF	124				
Db	59	RTKHIVDSAVRGFGTGFAYGQTNSGKTYTMWGSEDLHGVIPTRAIHDIQKIKFPDREF	118				


```
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 62672
; LENGTH: 955
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3689-241-A11_FLI.pcp
; US-10-425-114-62672
```

```
Query Match      28.8%; Score 714.5; DB 12; Length 955;
Best Local Similarity 39.6%; Pred. No. 4.3e-52;
Matches 193; Conservative 69; Mismatches 162; Indels 63; Gaps 15;

QY 1 AEEGAVACVVRPLNSREBSLGETAQVYWK--TDNNVIYQVDGS-----KSFNFRVVFH 53
Db 33 AKEKIFVTVRPLSKKE--LAAKDVAWECAQTQILYKGAQDRAAPMSYTFKVF 90
QY 54 GNETTKNVIYBEIAAPIIDSIAIQYNGTIFAYGQTASGTYTMMGSEDLHGVIPRAIHDF 113
Db 91 PACQTDVVEGAKDVAMSALTGINATIFAYGQTSSGKTFMR-----GVTSAVSDIY 144
QY 114 QTKKFPDRFLRVSYMEIYNETITDLLCGTQKMKPLIREDVNRNVIYVADLTEEVYIT 173
Db 145 RHIDSTPEREFVIKISAMEIYNETIVKDLL--RPDSAPFLRLDDPEKGTIVEKLOEEAKD 202
QY 174 SEMALKWITGKSRHYGETMKQRRSSHTIFRMILLESREKGPSPNCEGSKVSHNLV 233
Db 203 SQHLRHUISICEBQQRQVEALNDTSRSHQILKLTLESRLR-EVSCVKSF-VANLNFV 260
QY 234 DLASERAAQTGAAGVRLKEGCNINRSILFQGVIKKLSQGVGGFINYRDSKLTILQ 293
Db 261 DLASERAAQTHAIGARLKEGCHINRSILTLTVIRKLSGKSGHIPYRDSKLTILQ 320
QY 294 SLGNPKTRICTTTP--VSFDETLTALQFASAKYMNTPYVNEVSTDEALLKRYKEI 351
Db 321 SLGNARTAITICTMSPALTHVEQSRNTLFFATCAKEVNTAKVNNVSDKQLVKHLOTEL 380
QY 352 MDLKKQLEEVSLERAQ--AMEKDQALQLE-EXDLQKQVONEKIENLTRLMTVSSSLT 408
Db 381 ARLESELRTPDRGSSDILMEKDKIRQMEIEIEELRKQRD-----NA 424
QY 409 QOELKAKRKRVT----W-----CL---GKI-----NKKMSNYADQNPITNI 445
Db 425 QSELEELRKKVHDQPGWNPFDSPQARKKCLTFSGSLEPSNKNKNMSSIRQSSTAPF 484
QY 446 TTKXHL 452
Db 485 KHEIRKL 491
```

```
RESULT 7
US-10-116-712-670
; Sequence 670, Application US/10116712
; Publication No. US20030194764A1
; GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Switzer, Ann
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.568
; CURRENT APPLICATION NUMBER: US/10/116,712
; CURRENT FILING DATE: 2002-04-07
; NUMBER OF SEQ ID NOS: 670
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 670
; LENGTH: 1232
```

```
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-116-712-670

Query Match      28.7%; Score 713.5; DB 14; Length 1232;
Best Local Similarity 35.1%; Pred. No. 7.7e-52;
Matches 182; Conservative 86; Mismatches 191; Indels 59; Gaps 11;

QY 6 VAVCVVRPLNSREESLG-ETAQVYWKTDNNVIYQVDGSKSFNDFRVFHNKTKNVEE 64
Db 10 VRVALRCRPLVPKEISEGQCMCLSFVPGEPQVVVGT--KSFTYDFVDFPSTQEVEFT 67
QY 65 IAAPIIDSIAIQYNGTIFAYGQTASGTYTMMG-----SEDLGVIPRAIHDFQKIK 117
Db 68 AVAPLIKGVFKGNATVLAYGQTSGKTYSMGGAYTAQENETVGVIPRVIOLLFKRID 127
QY 118 KFPDRFLRVSYMEIYNETITDLLCGTQKMKPLIREDVNRNVIYVADLTEEVYITSEMA 177
Db 128 KXSDFTLKVSYLEIYNEIILLDLCPSEKAGINIREDPKEGKIVGLTEKTVLVALDT 187
QY 178 LKWITTKGKSRHYGETMKQRRSSHTIFRMILLESREKGPSPNCEGSKVSHNLVLDLAG 237
Db 188 VSCLEQGNNSRTVASTAMNSQSSRSASHAFTISLEQKKSD---KNSFSRSLHLVDLAG 243
QY 238 SERAAQTGAAGVRLKEGCNINRSILFQGVIKKLSQGVGGFINYRDSKLTILQNSLGG 297
Db 244 SERQKTKAEGDKLKEGININRGLLCLGNVISALGDDKGGFVYRDSKLTILQDSLGG 303
QY 298 NPXTRICTTTPV--SFDETLTALQFASAKYMNTPYVNEVSTDEALLKRYKEIMDLK 355
Db 304 NSHTLMACVSPADSNLEETLNLTVADRARKIKNPIVN-----IDPQTAELNHLK 355
QY 356 KQLEEVSL-----ETRAQAMEKDQALQLEEKDOLLQKQVONEKIENL 396
Db 356 QQVQQLQVLLQAHGGTLPGSITVPESENQSLMEKNQ--SLVEENEKLSRGLSEAAQGT 413
QY 397 TRML--VTSSSLTLOELKAKRKRVTWCLGKINK---MKSNYADQNPITNITKT 449
Db 414 AQMLERIILTEQANEKKNKXLEELRQHAACKLDLQKLVETLEQDELKENVEIICNLOOLI 473
QY 450 HKLSINLLRIBDSVCSSESVFSNTLDTLSEIENNPAT 487
Db 474 TQLS-----DETACMAAIDTAVQEQAQVETSPET 504
```

```
RESULT 8
US-10-408-765A-2153
; Sequence 2153, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2153
; LENGTH: 1232
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-408-765A-2153
```

```
Query Match      28.7%; Score 713.5; DB 16; Length 1232;
Best Local Similarity 35.1%; Pred. No. 7.7e-52;
Matches 182; Conservative 86; Mismatches 191; Indels 59; Gaps 11;
```


Db 143 HIENPTPERDPIIKISAMEIYNEIWKOLL--RPESTNLRLDDPEKGTIVKLEEBEIAKDS 200
Qy 175 EMALKWITKGEKSRHVGETHQWORSRSHTPIFRMILESEKREKGPSPNCEGSKVYSHNLVD 234
Db 201 OHRLHLISICEORQVGETALNDTSSSHQIILITVESRLR--EVSGCVKSF--VANLNFVD 258
Qy 235 LAGSERAQAOTGAAGVRLKEGCNINRSFILGQVKKLSGQVGFNYRDSKLTILQNS 294
Db 259 LAGSERAQAOTHAVGARLKEGCNINRSLLTTLVIRKLSDDKRSCHIPYRDSKLTILQLS 318
Qy 295 LGGNPKTRIICTITPVF---DETLTALQFASAKYMKNTPYNEVSTDEALLKRYKEI 351
Db 319 LGGNARTALICTMSPAQTHVEQSENITXLFATCAKEVTNNAKVMVYSDQLVKHLQMEV 378
Qy 352 MDLKKQLEEVSLTRAQ--AMEKDOLAQLLEKDL--LOKVON---EKIENLTRL--- 400
Db 379 ARLEAELRTPDRASSIIIMERDKIROQVEKEMELKQORDNAQSKLEELQKMGDNQ 438
Qy 401 -----VTSSSLTLOQELKAKRRRV 420
Db 439 PGWNPFDSPQTRKCLTYSGSLQSPNKNKI 468

RESULT 11
US-10-311-642-2
; Sequence 2, Application US/10311642
; Publication No. US20040086878A1
; GENERAL INFORMATION:
; APPLICANT: Cytokinetics, Inc.
; APPLICANT: Beraud, Christophe
; APPLICANT: Freedman, Richard
; TITLE OF INVENTION: NOVEL MOTOR PROTEINS AND METHODS FOR THEIR
; FILE REFERENCE: 020552-001910US
; CURRENT APPLICATION NUMBER: US/10/311,642
; CURRENT FILING DATE: 2003-09-29
; PRIOR APPLICATION NUMBER: 09/597,602
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1029
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: HsKif17 amino acid sequence
; OTHER INFORMATION: Description of Artificial Sequence: HsKif17
US-10-311-642-2

Query Match 28.6%; Score 710.5; DB 16; Length 1029;
Best Local Similarity 36.5%; Pred. No. 1.1e-51;
Matches 190; Conservative 80; Mismatches 176; Indels 75; Gaps 16;

Qy 5 AVAVCVRVRLNSREESLGETAQVYKTD-----NNVYQVDSKSFNDRVFHNE 56
Db 5 AVKVVVRCRPNQREEL--RCQPWTVVDCARACQCIQNPGADEPPKQFTFDGAYHVDH 62

Qy 57 TTKNVYVEIAPIDSALQGYNGTIFAYGQTASGKTYTMMGSED--HLGVIPIRAIHDF 113
Db 63 VTEQIYNEIAYPLVEGTEGNGTIFAYGQTSGSKFTMQGLPDPDSORGHIIPRAFHFV 122

Qy 114 QKIKKFPDRFLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNRYVADLTTEVYVT 173
Db 123 ESVQCAENTKFLVRASVLEIYNEVDRLDGADTKQK--LELKEHPEKGVYVVKLSMHTVHS 181

Qy 174 SEMALKWITKGEKSRHVGETHQWORSRSHTPIFRMILESR---EKGEPSNCEGSKVYSHL 230
Db 182 VAQCEHINETGWKRSVGYTLNMDKSDRSHTSIFTSIEMSAVDERG-----KDLRACKL 236

Qy 231 NLVDLAGSERAQAOTGAAGVRLKEGCNINRSFILGQVKKLSGQVGFNYRDSKLTIRI 290
Db 237 NLVDLAGSROSKTGATGERLKEATKINLSLSALGNVISALVDGRC--KHVYRDSKLTRL 295

Qy 291 LQNSLGNPKTRIICTITPV--SFDETLTALQFASAKYMKNTPYNEVSTDEALLKRYR 348
Db 296 LQDSLGNGTKTLMVACLSPADNNYDETLSTRYANRAKNIRKPRINEDPKD--ALLREYQ 354
Qy 349 KEIMDLKKQLEEVSLTRAQAMEKDOLAQL---BEKOLLQKV----- 388
Db 355 EEIKKKAIL-----TQOMSPSSLSALLSRQVPPDPVQVEKLLPQVIOHMEAEK 406
Qy 389 ---ONEKIENLTRLVTSSSLTLOQELKAKRRRVTCGLGKINKMKNSYADQFNIPNTNI 445
Db 407 QLRREEVEERLARL---KADYKAEQSSRARLEEDIT-----AMRNSYDVRSLTLEENL 456
Qy 446 TTKTHKLSINLREIDESVCSDDVFSNTLDTLSEIENPA 486
Db 457 RKETEAV-----LQGVLYKAEVMSRA--EFASSAEYPPA 489

RESULT 12
US-10-116-712-664
; Sequence 664, Application US/10116712
; Publication No. US20030194764A1
; GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Switzer, Ann
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.568
; CURRENT APPLICATION NUMBER: US/10/116,712
; CURRENT FILING DATE: 2002-04-07
; NUMBER OF SEQ ID NOS: 670
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 664
; LENGTH: 1232
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-116-712-664

Query Match 28.3%; Score 702.5; DB 14; Length 1232;
Best Local Similarity 35.1%; Pred. No. 6.8e-51;
Matches 182; Conservative 86; Mismatches 191; Indels 59; Gaps 12;

Qy 6 VAVCVRVRLNSREESLG-ETAQVYKTDNNVYQVDSKSFNDRVFHNETTKNVYEE 64
Db 10 VRVALRCRPLVPKEISEGCMCLSFVGPGEQVVVGTD--KSFTYDFVDFSTBOEVEVT 67
Qy 65 IAAPIIDSATQGYNGTIFAYGQTASGKTYTMMG-----SEHLGVIPIRAIHDFOKIK 117
Db 68 AVAPLIKGVFKGYNATVAYGQTSGKTYSMGGAYTAEQENEPTGVIPRVQLLFKRID 127

Qy 118 KFPDRFLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNRYVADLTTEVYVTSEMA 177
Db 128 KKSDFEFTLVKSVLEIYNEBILDLLCPREKAQINIREDPKGIKIVGLTEKTVLVALDT 187

Qy 178 LKMITKGEKSRHVGETHQWORSRSHTPIFRMILESEKREKGPSPNCEGSKVYSHNLVDLAG 237
Db 188 VSCLEQNNKRTVASTANVSQSSRSHAFTI----SLEQKKSKDNKSNFR--SKLHLVDLAG 243

Qy 238 SERAAQGAAGVRLKEGCNINRSFILGQVKKLSGQVGFNYRDSKLTILQNSLGG 297
Db 244 SEROKTKABEGRDLKEGINRGLLCLGNVISALGDDKKGGFAYRDSKLTIRLQDSLGG 303

Qy 298 NPKTRIICTITPV--SFDETLTALQFASAKYMKNTPYNEVSTDEALLKRYKEIMDLK 355
Db 304 NSHTLMIAVCSPADSNLEETLTLRYADRAKIKNKPIVN-----IDPQTAEHLNHLK 355

Qy 356 KQLEEVSL-----ETRAQAMEKDOLAQLLEKOLLQKVQNEKIENL 396
Db 356 QQVOOLQVILLQAHGGTLPDSITVEPSENLOSIMEKQ--SLVEENKLSGLSEAAQOT 413

Qy 397 TRML---VTSSSLTLOQELKAKRRRVTCGLGKINK---MKNSYADQFNIPNTITKT 449
Db 414 AQMLERIITWEOAKENKNAKLEELRQHAACKLDLQKLVTLEDQGLKENVEIICNLQOLI 473

Qy	450	HKLSINLLREIDESVCSSESDVFSNTLDTLSEIENPAT	487
Db	474	TQLS-----DETACMAAAIDTAVEQEQAQVETSPET	504

RESULT 13

```

US-10-116-712-669
; Sequence 669, Application US/10116712
; Publication No. US20030194764A1
; GENERAL INFORMATION:
;
; APPLICANT: Bangur, Chaitanya S.
;
; APPLICANT: Switzer, Ann
;
; TITLE OF INVENTION: COMPOSITIONS AND METHODS
;
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG
;
; FILE REFERENCE: 21021.568
;
; CURRENT APPLICATION NUMBER: US/10/116,712
;
; CURRENT FILING DATE: 2002-04-07
;
; NUMBER OF SEQ ID NOS: 670
;
; SOFTWARE: FastSeq for Windows Version 4.0
;
; SEQ ID NO 669
;
; LENGTH: 1232
;
; TYPE: PRT
;
; ORGANISM: Homo sapiens
US-10-116-712-669

```

Query Match	28.3%;	Score	702.5;	DB 14;	Length	1232;
Best Local Similarity	35.1%;	Pred. No.	6.8e-51;			
Matches 182;	Conservative	86;	Mismatches	191;	Indels	59;
					Gaps	12;

```

6  VAVCVVRPLNSRBSLG-BTAQVYWKTDNNVIYQVDSKSPFNDRVPHGNETTKNYEE 64
    QY
10 VRVALRCRPLVPKEISBGCQMLCSFVPGEPQVVVGTD--KSFTYDFVDPSTEQEEVNT 67
    Db
65 IAAPIIDSALQYNGNIFAFYQGTASGKTYTMMG-----SEDHLGVIPIRAIHDFOKIK 117
    QY
68 AVAPLLKGWFGYNATVLAYQGTGSGKTYSMGGAYTAEQENEPVTGVIPIVQILLFKEID 127
    Db
118 KFPDRFLLRVSYMEIYNETITDILCCTQKMKPLIREDVNRNVVYVADLTTEEVVYVTSMA 177
    QY
128 KXSDFEFTKVSVLEIYNEEILDLDCPSRQAQINREDPKGKIKIVGLGTEKTVLVALDT 187
    Db
178 LKWITKGEKSRHYGETKMNORSRSHTIFEMILESREKGPSPNCEGSKVYSHNLNVLDA 237
    QY
188 VSCLEGGNNSRTVASTAMNSQSSRSHAFTI---SLEQKKSDKNSSFR-SKULHVLDA 243
    Db
238 SERAATGAGVRLKGCNCINRSLFILGVQIKKLDGQVGGFPIRYRDSKUTRILQNSLGG 297
    QY
244 SERQKTKAEGDLKGINRGLLCLGNIVISALGDDKGGFAPYRDSKUTRILQDSLGG 303
    Db
298 NPKTRIICTIPV--SFEDETLTALQASTAKYMKNTPYNEVSTDEALLKRYKKEIMDLK 355
    QY
304 NSHTLMIACVSPADSNLEETLNTLRVADRARKIKNKPIVN-----IDPQTAELNHLK 355
    Db
356 KOLEEVS-----ETRAQAMEKDQLAQLLEEKDLOKQVNEKIENL 396
    QY
356 QQVQQLQVLLLOAHGGTLPGSITVPSENIQLSMEKNQ--SLVEENEKLSRGISEAQGT 413
    Db
397 TRML-----VTSSSLTLOEELKAKKRVRVTCWLGKINK---MKNSNYADQFNIPNTITKT 449
    QY
414 AQMLERIIWTEQANEXMNAKLEELRHOAAKCLDLQKLVLELQELKENVEIICNLQOLI 473
    Db
450 HKLSINLLREIDBSVCSESDVFNTLDLTSEIENPAT 487
    QY
474 TQLS-----DSTVACMAAAIDTAVEQEAQVETSPET 504
    Db

```

RESULT 14

US-10-437-963-169150
; Sequence 169150, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.

```

: APPLICANT: Zhou, Yihua
: APPLICANT: Cao, Yongwei
: APPLICANT: Wu, Wei
: APPLICANT: Boukharov, Andrey A.
: APPLICANT: Barbazuk, Brad
: APPLICANT: Li, Ping
: TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
: TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
: FILE REFERENCE: 38-211532211B
: CURRENT APPLICATION NUMBER: US/10/437,963
: CURRENT FILING DATE: 2003-05-14
: NUMBER OF SEQ ID NOS: 204966
: SEQ ID NO 169150
: LENGTH: 965
: TYPE: PRT
: ORGANISM: Oryza sativa
: FEATURE:
: OTHER INFORMATION: Clone ID: PAT_MRT4530_67597C.1.pap
US-10-437-963-169150

```

Query Match	27.8%	Score 691.5;	DB 16;	Length 965;
Best Local Similarity	40.8%	Pred. No. 4.2e-50;		

[illegible]

RESULT 15

```

US-10-408-765A-1664
; Sequence 1664, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1664

```



```

; LENGTH: 672
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-1664

Query Match      27.7%; Score 689; DB 16; Length 672;
Best Local Similarity 35.8%; Pred. No. 4e-50;
Matches 182; Conservative 81; Mismatches 151; Indels 94; Gaps 13;

Qy 31 KTDNNVIYQVDSKSNFDFVHGNETKNNVYEEIAAIIIDSAIOGNGTIFAYGQTASG 90
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 20 KTDSS---NEPPKTFFTDFVPESKQLDVNLTARPIIDSVLEGYNGTIFAYGQTG 75
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 91 KTYTMGSE---DHLGVIPIRAIHDIPOKIKKFP-DREFLLRVSMYIYNETITDLLCGTQ 146
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 76 KTFMEGVRAIPELRGIIIPNSPAHIFGHIKAEGRDFLRVSVLEIYNEEVRDLL-GKD 134
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 147 KMKPLIIREDVNRNVYVADLTEEVYVTSMALKWKTKGKSRHYGETKMNQRSRSHTIF 206
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 135 QTORLEVKERPVGVIYIKDLSAYVNNADDMRIMTLGHKNRSVGATNMNEHSSRSHAIF 194
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 207 RMILESEKGEPSCEGSKVSHLNVDLAGSEAAOTGAAGVELKEGCNINRSLFLGQ 266
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 195 TITIECEKGIDGNMH--VRWGKLUHVDLAGSERQAKTGQRLKEATKINLSLTLGN 252
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 267 VIKKLSQGVGGFINYRDSKLTIRLQNSLGNPKTRIITITPV--SFDETLTALQFAS 324
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 253 VISALVDGK-STHVPYRNSKLTIRLQDSLGNSTKMMCANIGPADYNYDETISTLRVANR 311
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 325 AKYMKNTPYVNEVSTDEALLKRYKEIMDLKKOLEE----- 360
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 312 AKNIKAKARINEDPKD-ALLFPQKETEELKKLEEGEISGSDISGSEEDDDDEGEVGE 370
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 361 -----VSLTRAQAMEKQQLALE---EKD 383
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 371 DGEKRRKRIQIGKKVSPDKMIENQAKIDEEKKALETKLDMEBEERNKAPAELEKREK 430
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 384 LLOKVQNEKIENLTRLMLVTSSSLTIQOELKAKRRRVTTWICIGKINKMKNSNYADQFNPT 443
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 431 LLKAAQOEHQ-----SLEKLSALEKKVI---VGGVDLLAKABEQEKLEES 473
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 444 NI-TTKTHKLSINLLREIDESVCSQSDV 470
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 474 NMELEERRKRAEQLRRELEKEKQERLDI 501
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Search completed: July 29, 2004, 10:06:29
Job time : 25.324 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 09:28:21 ; Search time 7.27641 Seconds
(without alignments)
6437.961 Million cell updates/sec

Title: US-10-045-631b-88_COPY_2_488

Perfect score: 2483
Sequence: 1 AEEGAVAVCVVRPLNSREE.....SDVFSNTDLTSLSEIWNPAT 487

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78:*
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2483	100.0	2563	1 S28261	centromere protein
2	1584	63.8	2954	2 T14156	kinesin-related pr
3	880	35.4	823	2 T52425	kinesin-like prote
4	844.5	34.0	888	2 D96619	protein T30E16.9 [
5	759	30.6	1459	2 T30196	kinesin motor prot
6	735.5	29.6	747	1 A57107	kinesin-related pr
7	728	29.3	701	1 B44259	kinesin-related pr
8	722	29.1	742	1 S58691	kinesin-related pr
9	716	28.8	786	2 A53939	kinesin homolog KH
10	710.5	28.6	1231	2 A54803	microtubule-associ
11	705	28.4	1121	2 T06065	hypothetical prote
12	704.5	28.4	699	1 S38982	kinesin-related pr
13	703	28.3	1066	1 A48669	kinesin-related pr
14	700	28.2	909	2 H86350	kinesin-related pr
15	699	28.2	1130	2 T21134	hypothetical prote
16	694	28.0	581	2 F84599	hypothetical prote
17	692	27.9	1226	2 I51617	probable kinesin h
18	689.5	27.8	932	2 T49235	kinesin-like prote
19	687.5	27.7	968	2 T45746	hypothetical prote
20	676.5	27.2	1225	2 A56514	Chromokinesin - ch
21	670	27.0	1263	2 T13465	hypothetical prote
22	664.5	26.8	1031	1 A38713	kinesin heavy chai
23	663.5	26.7	1254	2 T16277	kinesin heavy chai
24	660.5	26.6	975	1 A31497	kinesin heavy chai
25	656.5	26.4	834	2 T06055	hypothetical prote
26	656	26.4	672	2 S54351	kinesin osm-3 - Ca
27	654.5	26.4	963	1 A41919	kinesin heavy chai
28	652.5	26.3	1076	2 B84687	probable kinesin-l
29	652	26.3	793	2 JCS831	kinesin-related pr

30	650	26.2	1921	2 T13827	kinesin-73 - fruit
31	642.5	25.9	967	1 A35075	kinesin heavy chai
32	638.5	25.7	1032	2 I38510	neuronal kinesin h
33	635.5	25.6	881	2 I84737	kinesin heavy chai
34	635.5	25.6	968	2 T51933	kinesin motor prot
35	633	25.5	784	1 A55236	kinesin-related pr
36	632.5	25.5	1056	1 G02157	kinesin-like spind
37	631	25.4	1027	2 S37711	kinesin heavy chai
38	630.5	25.4	1067	2 S33417	kinesin-like prote
39	629	25.3	1060	1 A40264	kinesin-related pr
40	628	25.3	843	2 S44868	kinesin heavy chai
41	628	25.3	1388	2 T30335	Klp2 protein - Afr
42	624.5	25.2	857	2 E84600	probable kinesin h
43	624.5	25.2	1576	2 T29237	hypothetical prote
44	623.5	25.1	554	2 T50118	kinesin-related pr
45	623.5	25.1	1058	2 T47525	kinesin-related pr

ALIGNMENTS

RESULT 1
S28261
centromere protein E - human
N/Alternate names: centromere 312K protein; kinesin-related protein CENP-E
C/Species: Homo sapiens (man)
C/Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 19-Jan-2001
C/Accession: S28261
R/Yen, T.J.; Li, G.; Schaar, B.T.; Szilak, I.; Cleveland, D.W.
Nature 359, 536-539, 1992
A/Title: CENP-E is a putative kinetochore motor that accumulates just before mitosis.
A/Reference number: S28261, MUID:93024922; PMID:1406971
A/Accession: S28261
A/Molecule type: mRNA
A/Residues: 1-2663 <YEN>
A/Cross-references: EMBL:Z15005; NID:g29864; PIDN:CAA78727.1; PID:g29865
C/Genetics:
A/Gene: GDB:CENPE
A/Cross-references: GDB:361164; OMIM:117143
A/Map position: 4q24-4q25
C/Superfamily: centromere protein E; kinesin motor domain homology
C/Keywords: ATP; coiled coil; microtubule binding; mitosis; nucleotide binding; P-loop
F/7-335/Domain: kinesin motor domain homology <KWOT>
F/86-93/Region: nucleotide-binding motif A (P-loop)
F/86-2183/Domain: coiled coil #status predicted <COI>
F/92/Binding site: ATP (Lys) #status predicted

Query Match		100.0%;	Score 2483;	DB 1;	Length 2663;
Best Local Similarity		100.0%;	Pred. No. 1.2e-134;		
Matches 487;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	AEEGAVAVCVVRPLNSREESLG	TAQVYWKTDNNVIYQV	DGSKGFNFDRVPHGNETTKN	60
Db	2	AEEGAVAVCVVRPLNSREESLG	TAQVYWKTDNNVIYQV	DGSKGFNFDRVPHGNETTKN	61
Qy	61	VYEETAAPLIDSAIOGYNGTIF	FAYGQTASGKTYTMGSE	DHLGVIPRAIHDIFQIKKPP	120
Db	62	VYEEAAPLIDSAIOGYNGTIF	FAYGQTASGKTYTMGSE	DHLGVIPRAIHDIFQIKKPP	121
Qy	121	DREFLLRYSYMEIYANETITD	LCCGQKMKPLIIRDVNR	NVYVADLTSEVYVTSMALAK	180
Db	122	DREFLLRYSYMEIYANETITD	LCCGQKMKPLIIRDVNR	NVYVADLTSEVYVTSMALAK	181
Qy	181	ITKGSKSHYGETKMNQSSRSHT	IFRMILSREKGEPSNCE	SGSVKVSHLNLVDLAGSR	240
Db	182	ITKGSKSHYGETKMNQSSRSHT	IFRMILSREKGEPSNCE	SGSVKVSHLNLVDLAGSR	241
Qy	241	AAQTCAAGVRLKEGCNINRSL	FILGOVVKLSGQGVGF	INRYRDSKLTILQNSLGGNPK	300
Db	242	AAQTCAAGVRLKEGCNINRSL	FILGOVVKLSGQGVGF	INRYRDSKLTILQNSLGGNPK	301
Qy	301	TRIICTITPVSFDETLTALQF	ASTAKYMKNTPYNEVST	DEALLKRYBKEIMDLKKOLEE	360

Db 302 TRIICITTPVSPDETLTALQFASAKVMKNTPYVNEVSTDEALLKRYRKEIMDLKKOLEE 361
Qy 361 VSLTRQAQMEKDLQALILEKDLLOKVQNEKIENLFRMLVTSSSLTQQLKAKRRRV 420
Db 362 VSLTRQAQMEKDLQALILEKDLLOKVQNEKIENLFRMLVTSSSLTQQLKAKRRRV 421
Qy 421 TWCLGKINKMNSYADQFNPTNITTKTKLSINLAREIDESVCSBDSVFNLTDLTSE 480
Db 422 TWCLGKINKMNSYADQFNPTNITTKTKLSINLAREIDESVCSBDSVFNLTDLTSE 481
Qy 481 IEWNPAT 487
Db 482 IEWNPAT 488

RESULT 2

Ti14156
kinesin-related protein - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-May-2000
C:Accession: T14156
R:Wood, K.W.; Sakowicz, R.; Goldstein, L.S.; Cleveland, D.W.
Cell 91, 357-366, 1997
A:Title: CENP-E is a plus end-directed kinetochore motor required for metaphase chromosome segregation
A:Reference number: Z17893; PMID:98028574; PMID:9363944
A:Accession: T14156
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2954 <WOO>
A:Cross-references: EMBL:AF027728; NID:G2586070; PID:G2586071; PIDN:AAC60300.1
C:Genetics:
A:Gene: XCENP-E
C:Superfamily: Centromere protein E; kinesin motor domain homology

Query Match 63.8%; Score 1584; DB 2; Length 2954;
Best Local Similarity 64.0%; Pred. No. 7.4e-83;
Matches 320; Conservative 68; Mismatches 96; Indels 16; Gaps 6;
Qy 1 AEEGAVCVVRPLNSREESIGETAQVYVKTNNVYQVDSKSFNDRVPHGNETTKN 60
Db 2 SEGDAVVCVVRPLNIREQ--GDOANLQWKAGNNTISQVDTGKSFNDRVPHGNETTKN 59
Qy 61 VVEEIAAPIIDSAIQNGYTFAYGQTASGKTYTMMGSEDLGVIPRAIHDFQIKKPP 120
Db 60 IYQELAVLIIRALQNGYTFAYGQTSSGKTYTMMGTPNSGLIPQAIQEVFKIIEIP 119
Qy 121 DREFLLRVSYMEIYNETITDLGCTQKMKPLIREDVNRVYVADLTVEVYVTSWALKN 180
Db 120 NREFLLRVSYMEIYNETITDLGCTQKMKPLIREDVNRVYVADLTVEVYVTSWALKN 179
Qy 181 ITGKSKRHYGTEKQNRSSRSHTIFRMLERKEGPE--NCEGSKVVKSHLNLVDLAG 237
Db 180 IKKGKSKRHYGTEKQNRSSRSHTIFRMLERKEGPE--NCEGSKVVKSHLNLVDLAG 239
Qy 238 SERAAQTGAAGVRLKEGCKNINSLFILGVIKKLSGQVGGFVINYRDSKLTILQNSLGG 297
Db 240 SERASQTGAAGVRLKEGCKNINSLFILGVIKKLSGQVGGFVINYRDSKLTILQNSLGG 299
Qy 298 NPKTRIICTITPVSPDETLTALQFASAKVMKNTPYVNEVSTDEALLKRYRKEIMDLKKQ 357
Db 300 NAKTVIICITITPVSPDETLTALQFASAKVMKNTPYVNEVSTDEALLKRYRKEIMDLKKQ 359
Qy 358 LE--SVSLTRQAQMEKDLQALILEKDLLOKVQNEKIENLFRMLVTSSSLTQQLKAK 415
Db 360 LENLSSSTKQAQMAKEHTQLAIEIKQLHKEKREDRIWHLNIIVASSQES--QDQRVK 418
Qy 416 RKRRTVCLGKINKMNSYADQFN----IPNITTKTKLSINLAREIDESVCSBDSVFN 471
Db 419 RKRRTVCLGKINKMNSYADQFN----IPNITTKTKLSINLAREIDESVCSBDSVFN 478
Qy 472 SNVLTDLTSL--EIEWNPAT 487
Db 479 DDALSMDSNGIDAENNLAS 498

RESULT 3

T52425
kinesin-like protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 24-Oct-2000 #sequence_revision 24-Oct-2000 #text_change 24-Oct-2000
C:Accession: T52425
R:Kato, A.; Suzuki, M.; Kuwahara, A.; Ooe, H.; Higano-Inaba, K.; Kameda, Y.
Gene 239, 309-316, 1999
A:Title: Isolation and analysis of cDNA within a 300 kb Arabidopsis thaliana genomic re
A:Reference number: Z25171
A:Accession: T52425
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-823 <KAT>
A:Cross-references: EMBL:AB028468; PIDN:BAAB8112.1
C:Genetics:
A:Gene: ZCF125

Query Match 35.4%; Score 880; DB 2; Length 823;
Best Local Similarity 43.1%; Pred. No. 4.4e-43;
Matches 220; Conservative 70; Mismatches 166; Indels 54; Gaps 12;
Qy 6 VAVCVVRPLNSREESIGETAQVYVKTNNVYQVDSKSFN-----FDRVPHGNE 56
Db 4 ICVAVVRP-----PAPENGASLWKVEDN--RISLHKSUDTPIITASHAPDHVDESS 54
Qy 57 TTKNVVEIAAPIIDSAIQNGYTFAYGQTASGKTYTMMGSEDLGVIPRAIHDFQIK 116
Db 55 TNASVYELLTKDIIHAAVEGFGNFAFGQTSSGKTFMTGSETDPGIIRSVRDPFRI 114
Qy 117 KFPDREFLLRVSYMEIYNETITDLGCTQKMKPLIREDVNRVYVADLTVEVYVTSW 176
Db 115 HWSIDREFLLRVSYMEIYNEBNDLL--AVENQRLOIHEHLERGVSFVAGLKEEIVSDAEQ 172
Qy 177 ALKMITGKSRHYGTEKQNRSSRSHTIFRMLERKEGPE--NCEGSKVVKSHLNLVDLA 236
Db 173 ILKLIDSGVNRHFGETNMNVHSSRSHTIFRMTIESR--GKDNSSSDAIRSVSLNLVDLA 230
Qy 237 GSERAAQTGAAGVRLKEGCKNINSLFILGVIKKLSGQVGGFVINYRDSKLTILQNSL 295
Db 231 GSERIAKTGAAGVRLQEGKYINKSLMILGNVINKLSDSTKLRAHIPYRDSKLTILQPAL 290
Qy 296 GGNPKTRIICTITPVSP--FDELTALQFASAKVMKNTPYVNEVSTDEALLKRYRKEIMD 353
Db 291 GGNAKTCTIICITIAPEHHIESKGLTQFASAKRITNCAQVNEILTDALLKROKLEIEE 350
Qy 354 LKKQLEEVSLTRAQ-----AMKNDQLAQLLEKDLLOKVQNE-----KIE 394
Db 351 LRMKLQGSAAEVLQOEILNLNQLMKYELCEERLTKQLEEEKRKEGCKNKEQOMKIE 410
Qy 395 NLTRMLVTSSSLTQQLKAKRRVTVCLGKINKMNSYADQFNIPNITTKTKLSI 454
Db 411 NLNN--FVINSDF-----KRNQSEDFIISKTPDGLCNVNDTSDVPTGTCFKSASRSF 461
Qy 455 NLLREIDESVCSBDSVFNLT--DTLSEIEW 483
Db 462 VVARSNNSYGLSDFSPVHSLGDAVEDTWT 491

RESULT 4

D96619
protein T30B16.9 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: D96619
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, C.; Alonso-
Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.
ansen, N.F.; Hughes, B.; Huijzer, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: D96619
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-888 <STO>
A:Cross-references: GB:AR005173; NID:9877739; PIDN:AAF79747.1; GSPDB:GN00141
C:Genetics:
A:Gene: T30E16.9
A:Map position: 1.

Query Match 34.0%; Score 844.5; DB 2; Length 888;
Best Local Similarity 40.0%; Pred. No. 5.4e-41;
Matches 220; Conservative 70; Mismatches 169; Indels 91; Gaps 13;
QY 6 VAVCVRPVPLNSREESLGETAQVYVTKTDNVI-----Y 38
DB 4 ICVAVVRP-----PAPENGASLWKVEDNRISLHKSILDTPTTASHAFVSGISITDLI 57
QY 39 QVDSKSFNF-----DRVPHGNETTKVYVEIAAIIIDSAIOGYNGTIFAYGQTA 88
DB 58 EIVSLFLFSGVYVFFLLFADHVFDESSNASVYELTLDKIIHAAVEGFNGTAFAYGQTS 117
QY 89 SGKTYTMGSEDLHGVIPRAIHDFQIKKFPDPREFLLRVSYMEIYNETITDILLCGTQKM 148
DB 118 SGKTYTMGSETDPIIRRSVRDVFERHMSIDREFLIRVSYMEIYNEEINDLL--AVEN 175
QY 149 KPLIIRVNVVADLVEVYVSEMALKWITKGEKSRHYGETKMNQSRSSHTIFRM 208
DB 176 QRLQIHEHLERGTVFVAGLKEEIVSDAEQILKIDSGEVNRHFGETNNVHSSRSHTIFRM 235
QY 209 -----ILESGKEPCNCEGSKVSHNLVLDAGSRAAQGTAGVRLKEGCN 256
DB 236 VPRFSYERDLLVIER--GKNSSSDAIRVSNLVNLDLAGSRIAKTGAGGVRLQEGKY 293
QY 257 INRSFILQVVKKLDG-QVGFNRYRDSKLTFRILQNSLGNPKPRTIITIPVS--PD 313
DB 294 INKSMILGNVINKLSDSKLRAHPIYRDSKLTFRILQALGNGNAKTCIITIAPEEHIE 353
QY 314 ETLTALQFASAKYKNTVYNEVSTDEALLKRYRKEIMDLKKEEVSLETRAQ----- 368
DB 354 ESKGTILQFASRAKRIITCAQVNEILDAALLKQKLEIEELRWKLGQSHAEVLEQELNL 413
QY 369 -----AMEKQOLAQLLEKDLQKQVNE-----KLENLRLMVTSSSLTLOQLKA 414
DB 414 SNQMLKYELECEBLKTLQLEEKKKQKEQNCIKBQMKIENLNN-FVTNSDF----- 464
QY 415 KKRVRVTCGLKINKMKNVADQFNPTNITTKTKHLSINLIRIDEISVCSDESDFVSN 474
DB 465 KRNQSEDFIISRKTPDGLCNVNDTSVPCTPCFKSASRFFVARSNNYSLGLDFSPMWS 524
QY 475 L-DTLSEITW 483
DB 525 LGDVADEDTW 534

RESULT 5
T30196
Kinesin motor protein 1 - smut fungus (Ustilago maydis)
C:Species: Ustilago maydis (corn smut)
C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 11-May-2000
C:Accession: T30196
R:Lehmler, C.; Steinberg, G.; Snettselaar, K.M.; Schliwa, M.; Kahmann, R.; Bolker, M.
A:Title: Identification of a motor protein required for filamentous growth in Ustilago
A:Reference number: Z20770; MUID:97361828; PMID:9218789
A:Accession: T30196
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1459 <LEH>

A:Cross-references: EMBL:U92844; NID:92062749; PID:92062750; PIDN:AAB63336.1
C:Genetics:
A:Gene: kin1
C:Function:
A:Description: required for filamentous growth in Ustilago maydis
Query Match 30.6%; Score 759; DB 2; Length 1459;
Best Local Similarity 38.2%; Pred. No. 8.7e-36;
Matches 191; Conservative 71; Mismatches 128; Indels 110; Gaps 14;
QY 6 VAVCVRPVPLNSREESLGETAQVYVTKTDN--NVLYQVD----- 41
DB 237 VVVCVRP-----SRASSSDSEASV-WNCDEKRIPTPEHHPAIAKRTTSSRAGAGASIA 293
QY 42 -----GSKSFNDFRVPHGNETTKVYVEIAAIIIDSAIOGYNGTIFAYGQ 86
DB 294 AAPSSHDLHEDPTSTYHFQDKLITGAQTDDMYHSHAPVRAAVEGNGTVFAYGQ 353
QY 87 TASCCTYTMGSEDLHGVIPRAIHDFQIKKFPDPREFLLRVSYMEIYNETITDILLCGTQ 146
DB 354 TSGKTHMTSGSDAEPGVIPRAVEQIFQMIKDEPREFLLRVSYMEIYNETLKDLLA--- 410
QY 147 KMKPL-----IREDVNVNVVADLVEEV 171
DB 411 PLPLTGTSGTSSLTQTRPASPPIKSGSSHAAGSQSCTLRILIEDQKSSRVIITGLREEIV 470
QY 172 YTEMLAKWITKGEKSRHYGETKMNQSRSSHTIFRMILESREKGEPCNCEGSKVSHLN 231
DB 471 TDANTVLCIQRQDERHVGATDWNERSRSHCVQLTIESRSPASKE--VRISQLN 528
QY 232 LVDLAGSRAAQGTAGVRLKEGCNINRSILFQGVVKLSDGQVG--FINYRDSKLT 289
DB 529 LIDLAGSRAA---SOAERKEGAFINKSLITGTVIGLTFEPVNGDAHIPYRDSKLT 585
QY 290 ILQNSLGNPKPRTIITIPVS--FDETLTALQFASAKYKNTVYNEVSTDEALLKRY 347
DB 586 ILQTSLSGNARIIVITLSPDTEHNETLSTLFGKRCKLVTITAKKGTAMDKALLQKY 645
QY 348 RKEIMDLKQLEB-----VSLETRAQAMEK-DQLAQLLE---EKDLQKQV 389
DB 646 RKELDALRAKLEANGSPNPNVMTIVSABASKESQKLDQLNQKEAAQREVEDMQKKR 705
QY 390 NE---KLENLRLMVTSSSL 406
DB 706 SHLKAQIEHLTRLILTQSIV 725
RESULT 6
A57107
Kinesin-related protein KIF3B - mouse
C:Species: Mus musculus (house mouse)
C>Date: 03-Nov-1995 #sequence_revision 03-Nov-1995 #text_change 19-Jan-2001
C:Accession: A57107
R:Yamazaki, H.; Nakata, T.; Okada, Y.; Hirokawa, N.
J. Cell Biol. 130, 1387-1399, 1995
A:Title: KIF3A/B: a heterodimeric kinesin superfamily protein that works as a microtubul
A:Reference number: A57107; MUID:96032268; PMID:7559760
A:Accession: A57107
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-747 <YAM>
A:Cross-references: GB:D26077; NID:gl060922; PIDN:BAA05070.1; PID:gl060923
A:Experimental source: brain
C:Complex: heterodimer with KIF3A (PIR:B44259); the KIF3A/3B heterodimer associates with
C:Function:
A:Description: KIF3 complex is a motor protein that provides anterograde axonal tra
C:Superfamily: kinesin-related protein KIF3; kinesin motor domain homology
C:Keywords: ATP; coiled coil; heterodimer; heterotrimer; microtubule binding; nucleotide
F:1-363/Domain: head globular #status predicted <HGL>
F:10-346/Domain: kinesin motor domain homology <KMOT>
F:96-103/Region: nucleotide-binding motif A (P-loop)
F:364-592/Domain: helical rod #status predicted <ROD>
F:594-747/Domain: tail globular #status predicted <TGL>

F;102/Binding site: ATP (Lys) #status predicted	
Query Match	29.6%; Score 735.5; DB 1; Length 747;
Best Local Similarity	37.9%; Pred. No. 8.1e-35;
Matches 193; Conservative	85; Mismatches 180; Indels 51; Gaps 13;
QY	5 AVAVCVRRPLNRSRESLG-----ETAQVYWKTDNNVIYQVDSGKSFNDRVPHGN 55
Db	9 SVRVVRCRPMNGKEKAASVDKVDVVKLGQVSVKPKGTSHM--PKTFTDAVVDWN 66
QY	56 ETTKNVYEELAAPIDSALIOYNGTIFAYGQTASGKTYTM---MGSEDLHGIVPRAIHDI 112
Db	67 AKQELVDEFRPLVDVLQGFNGTIFAYGQTGKTYTMEGVGRGDPKRGVIPSFDHI 126
QY	113 FQIKKPPDRFLRVSYMEIYNTITDLCGTQKMKPLIREDVNRNVVADLTEEVDY 172
Db	127 FTHISRQOQYLVRASYLEIYQEIIRDLSSKDQ--TKRLERKRPDTGVVYVKDLSSFTVK 185
QY	173 TSEMALKWITKESRHYGETKMNQSRSSHTIFRMILLESKEGPNCEGSKVSHLNL 232
Db	186 SVKIEHVMNVGNQNRSGVATNMHESSRSRAIFVITIECSEVG--LDGENHIRVGKLN 243
QY	233 VDLAGSRAAQTGAAGVRLKEGCNINRSFLILGVQVKKLSGQVGGFFINYSRDKLTRILO 292
Db	244 VDLAGSERQAKTGAGGERLKEATKINLSLGNVISALVDGK--STHIPPYRDSKLTRILO 302
QY	293 NSLGNPKTRICTITPVSF--DETILTAQFASAKYMKNTPYVNEVSTDEALLKRYRKE 350
Db	303 DSLGNGAKTVMVANVPASYNVEETLTILRYANRAKNIKPRVNEDEPKD--ALLREFORE 361
QY	351 IMDLKKOLEEVSLETRAQ-----AMEKDOQLAELEE---KDLQKQVQNEKI 393
Db	362 IARLKAQLEKRSIGRRKREKRGSGGGGEEEGEEDGDDKDDVWRQEQEKL 421
QY	394 ENLTRLMLVTSLSLTLLQELK-----AKRKRVTWCLG-KINKQNSNYADQFNI 441
Db	422 ELEKRAIVEDHSLVAEERMLKEXEKMEDLRREKDAEMLGAKIKAMESKLLVGGKNI 481
QY	442 PNIITKTHKLSINLLREIDESVCSSESDV 470
Db	482 -VDHTNEQOKILEQRKQETAEQKRREI 509
RESULT 7	
B44259	
Kinesin-related protein KIF3A - mouse	
C:Species: Mus musculus (house mouse)	
C:Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 19-Jan-2001	
C:Accession: B44259; S27872	
R:Alzawa, H.; Sekine, Y.; Takemura, R.; Zhang, Z.; Nangaku, M.; Hirokawa, N.	
J. Cell Biol. 119, 1287-1296, 1992	
A:Title: Kinesin family in murine central nervous system.	
A:Reference number: A44259; MUID:93077686; PMID:1447303	
A:Accession: B44259	
A:Molecule type: mRNA	
A:Residues: 1-701 <AIZ>	
A:Cross-references: EMBL:D12645; NID:g220469; PIDN:BAA02166.1; PID:g220470	
A:Experimental source: brain	
A>Note: sequence extracted from NCBI backbone (NCBIP:118911)	
C:Complex: heterodimer with KIF3B (PIR:A57107); the KIF3A/3B heterodimer associates with	
C:Function:	
A:Description: KIF3 complex is a motor protein that provides anterograde fast axonal tra	
C:Superfamily: kinesin-related protein KIF3; kinesin motor domain homology	
C:Keywords: ATP; coiled coil; heterodimer; heterotrimer; microtubule binding; nucleotide	
F;1-368/Domain: head globular #status predicted <HGL>	
F;15-351/Domain: kinesin motor domain homology <KMT>	
F;100-107/Region: nucleotide-binding motif A (P-loop)	
F;369-599/Domain: helical rod #status predicted <ROD>	
F;600-701/Domain: tail globular #status predicted <TGL>	
F;106/Binding site: ATP (Lys) #status predicted	
Query Match	29.3%; Score 728; DB 1; Length 701;
Best Local Similarity	41.7%; Pred. No. 2e-34;
Matches 198; Conservative	74; Mismatches 191; Indels 64; Gaps 13;
QY	6 VAVCVRRPLNRSRESLGETAQVYWKTDNNVIYQVDSGKSFNDRVPHGNETT 58
Db	9 KVYVRCRPMNSKEISQGHKRVEMDNKRLVEVTNPKGPPGPNKSFDTVDVWNSKQ 68
QY	59 KNYVEIAAPIDISALIOYNGTIFAYGQTASGKTYTMG---SEDLHGIVPRAIHDFQK 115
Db	69 IDLYDEFRSIVSVLQGFNGTIFAYGQTGKTYTMEGVRSNPELGRVIPSNEHIFTH 128
QY	116 IKKPPDRFLRVSYMEIYNTITDLCGTQKMKPLIREDVNRNVVADLTEEVDY 175
Db	129 IARTNQOFLVRASYLEIYQEIIRDLAKDQK-KRLDLKERPDGTGVYVKDLSSFTVSKV 187
RESULT 8	
S58691	
Kinesin-related protein KRP95 - sea urchin (Strongylocentrotus droebachiensis)	
N:Alternate names: kinesin-2 chain B; KRP (85/95) 95K chain	
C:Species: Strongylocentrotus droebachiensis	
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-Jan-2001	
C:Accession: S58691	
R:Rashid, D.J.; Wedaman, K.P.; Scholey, J.M.	
J. Mol. Biol. 252, 157-162, 1995	
A:Title: Heterodimerization of the two motor subunits of the heterotrimeric kinesin, KRP95	
A:Reference number: S58691; MUID:95404610; PMID:7674298	
A:Accession: S58691	
A>Status: preliminary; nucleic acid sequence not shown; not compared with conceptual	
A:Molecule type: mRNA	
A:Residues: 1-742 <RAS>	
C:Complex: heterotrimer of a 115K chain and two kinesin-related chains of 85K (PIR:S38959)	
C:Superfamily: kinesin-related protein KIF3; kinesin motor domain homology	
C:Keywords: ATP; heterotrimer; microtubule binding; nucleotide binding; P-loop	
F;9-345/Domain: kinesin motor domain homology <KMT>	
F;95-102/Region: nucleotide-binding motif A (P-loop)	
F;101/Binding site: ATP (Lys) #status predicted	
Query Match	29.1%; Score 722; DB 1; Length 742;
Best Local Similarity	37.6%; Pred. No. 4.8e-34;
Matches 198; Conservative	74; Mismatches 191; Indels 64; Gaps 13;
QY	6 VAVCVRRPLNRSRESLGETAQVYWKTDNNVIYQVDSGKSFNDRVPHGNETT 58
Db	9 KVYVRCRPMNSKEISQGHKRVEMDNKRLVEVTNPKGPPGPNKSFDTVDVWNSKQ 68
QY	59 KNYVEIAAPIDISALIOYNGTIFAYGQTASGKTYTMG---SEDLHGIVPRAIHDFQK 115
Db	69 IDLYDEFRSIVSVLQGFNGTIFAYGQTGKTYTMEGVRSNPELGRVIPSNEHIFTH 128
QY	116 IKKPPDRFLRVSYMEIYNTITDLCGTQKMKPLIREDVNRNVVADLTEEVDY 175
Db	129 IARTNQOFLVRASYLEIYQEIIRDLAKDQK-KRLDLKERPDGTGVYVKDLSSFTVSKV 187

355 KQLEEVSL-----ETRAQAMEKQDLAQLEEKDLQKQVNEKIE 395
 356 KQVQOQIQLLQAHGTLPGDINVEPSENLOSLMEKQ--SLVEENEKLSRGLSEAAQ 413
 396 LTRML---VTSSLTLOELKAKRKRVTVCLGKINK-----MKNSYADQFNPTNTTK 448
 414 TAQMLERIIITEQANERKNKLEELRRHAACKVDLQKLVETLEDELKENIEIICNQV 473
 449 THKLS-----INLLRBIDSVCS-----ESDVFS 472
 474 IAQLSDEAAACMTATIDTAGEDATQVQSSPDTGRSSDVFS 513

RESULT 11
 T06065
 Hypothetical protein F19H22.150 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 31-Mar-2000
 C:Accession: T06065
 R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X
 submitted to the Protein Sequence Database, March 1999
 A:Reference number: Z15184
 A:Accession: T06065
 A:Molecule type: DNA
 A:Residues: 1-1121 <BEV>
 A:Cross-references: EMBL:AL035679; GSPDB:CN00062; ATSP:F19H22.150
 A:Experimental source: cultivar Columbia; BAC clone F19H22
 C:Genetics:
 A:Gene: ATSP:F19H22.150
 A:Map position: 4
 A:Introns: 139/2; 170/1; 200/1; 256/3; 284/3; 302/3; 322/3; 349/1; 371/3; 390/3; 425/3;
 C:Superfamily: kinesin heavy chain; kinesin motor domain homology
 F:99-494/Domain: kinesin motor domain homology <KMOT>

Query Match 28.4%; Score 705; DB 2; Length 1121;
 Best Local Similarity 34.3%; Pred. No. 7.8e-33;
 Matches 196; Conservative 96; Mismatches 148; Indels 132; Gaps 18;

QY 1 AEEGAVCVVRPLN-----SREES-----LGETA---QYVWKTDN 34
 DB 94 SERDSISVTRFRLYARSDLAMDQYICRDYSFHDVAIGVNSLLGQDSFGIPLAIDVF 153
 QY 35 NVI--YQVDSKSFNFRVPHGNETTKNVYEIAAIDSAIQYNGTIFAYGOTASGKT 92
 DB 154 DTLVREXNPLTAAFDKVGQPGQATIDVDVAPPVKAAWEGVNGVFAVGYTSSGKT 213
 QY 93 YTM-----MGSEDLHLGVPRAIHDIF 113
 DB 214 HTMHRVRLKKNPYTPFSLVFLQNTCAMNWDLFNLVPLWYLLGQDSFGIPLAIDVF 273
 QY 114 QKIKF-----PDREFLRYVMEIYNETITDLCGQTKMKPLIREDVNRNVVADLT 167
 DB 274 SIIDQVSLGNTPGREFLRYVMEIYNETITDLCGQTKMKPLIREDVNRNVVADLT 330
 QY 168 BEVYVTSMLAKWITKGEKSRHYGETKQNRSSRSHITFRMILSRKGEPSNCEGSYKV 227
 DB 331 BEVVLSPGHLSFIAAGEEHRHVGSNFNLLSSRSHITFTLWESSATGDEYD---GVIF 387
 QY 228 SHLNIIDLAGEAAQAGVRLKEGCNINRSIFILQVTKKLSGQGVGFYNYRDSKL 287
 DB 388 SQLNIDLAGSE--SSKTETTLGRKEGSIYKNSLLTLGTIVGLSEK--ATHIYRDSKL 445
 QY 288 TRILONSIGGNPKTRITCTITPV--SPDETTLAQFASTAKYMKNTPVYNEVSTDEALLK 345
 DB 446 TRLQSSLSHGHSGLICTIITPASSSEETNLKFAKRAKSIETIYASNQIIEKSLIK 505
 QY 346 RYRKEIMDLKKQLEEV-----SLETRAQAMEKDQ--LAQLLEEKOLLQKQVNEK 392
 DB 506 KYQREISTKLDELQRLRGMLVGVSHEELMSLKQLEEGQVKMQSRLEEEBAKALMSR 565
 QY 393 IENLTRMLVTSLSLTLOELKAKRKRVTVCLGKINKMKNYADQFNPTNTTKTKL 452

566 IQKTLKILVST-----KNSIFQYSGDIP-----THQR 593
 QY 453 SINLLREID--EVSVCSESDVF---SNTLDTLSE 480
 DB 594 SLSAGKDRFDSLLLESNDNLGSPSSITALLSE 625

RESULT 12
 S38982
 kinesin-related protein KRP85 - sea urchin (Strongylocentrotus purpuratus)
 N:Alternate names: kinesin-2 chain A; KRP (85/95) 85K chain
 C:Species: Strongylocentrotus purpuratus (purple urchin)
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-Jan-2001
 C:Accession: S38982; S72551
 R:Cole, D.G.; Chinn, S.W.; Wedaman, K.P.; Hall, K.; Vuong, T.; Scholey, J.M.
 Nature 366, 268-270, 1993
 A:Title: Novel heterotrimeric kinesin-related protein purified from sea urchin eggs.
 A:Reference number: S38982; MUID:94050179; PMID:8222586
 A:Molecule type: mRNA
 A:Accession: S38982
 A:Residues: 1-699 <COL1>
 A:Cross-references: EMBL:L16993; NID:g295245; PIDN:AAA16098.1; PID:g295246
 A:Accession: S72551
 A:Molecule type: protein
 A:Residues: 2-5, 'X', 7-11; 59-64; 125-132; 222-226, 'X', 228-230 <COL2>
 C:Complex: heterotrimer of a 115K chain and two kinesin-related chains of 95K (PIR:S586)
 C:Superfamily: kinesin-related protein KIF3; kinesin motor domain homology
 C:Keywords: Atp; heterotrimer; microtubule binding; nucleotide binding; P-loop
 F:11-348/Domain: kinesin motor domain homology <KMOT>
 F:97-104/Region: nucleotide-binding motif A (P-loop)
 F:103/Binding site: ATP (Lys) #status predicted

Query Match 38.4%; Score 704.5; DB 1; Length 699;
 Best Local Similarity 36.2%; Pred. No. 4.5e-33;
 Matches 200; Conservative 76; Mismatches 152; Indels 125; Gaps 19;

QY 6 VAVCVRVPLNSREESLG-----ETAQYVWKTDNNVYQVDSKSFNFRVPHG 54
 DB 11 VRVVRCPRLSKETGQFKSVVXMDMRGTQV---TNPNA-PSGPPPKSFTFDTVPAP 66
 QY 55 NEITKNVYEIAAIDSAIQYNGTIFAYGOTASGKTYYMMG--SEDHL-GVIPRAIHD 111
 DB 67 GAKQTDVYNQATRPIDVDAILEGNTIFAYGOTGKTGKTFTMEGVRSGPELGIPLNSFAH 126
 QY 112 IFQKIKFPDR--EFLRVSYMEIYNETITDLCGQTKMKPLIREDVNRNVVADLTVEV 170
 DB 127 IFGHIAKEQENVRFLRVSVLEIYNEEVKDLL-GKDQQRHLEVKERPDVGVYVKDLGSAFV 185
 QY 171 VYTSMLAKWITKGEKSRHYGETKQNRSSRSHITFRMILSRKGEPSNCEGSYKVSHL 230
 DB 186 VNNADDMDRIMTLGNKRSVGNATNMNSESRSRSHITFTLERSDMG--LDKEQHRVRGKL 243
 QY 231 NLVDLAGSRAAQAGVRLKEGCNINRSIFILQVTKKLSGQGVGFYNYRDSKLTRI 290
 DB 244 HMVDLASERQTKTGATGQRLKEATKLNLSLTGLNVLSSLDVCK-SHTIYRNSKLTRL 302
 QY 291 LQNSLGNPKTRITCTITPV--SPDETTLAQFASTAKYMKNTPVYNEVSTDEALLKRYR 348
 DB 303 LQDSLGNVAKTVMCANIGPAEYNYDETISTLRVANRAKNIKAKINAKINEDPKD-ALLRBFQ 361
 QY 349 KEIMDLKKQLEEV-----VSLTRAQAM 370
 DB 362 KEIEELKQISESGLEGLDDDESGSESGDEEAGEGVKKRKRKKNPKRKLPEIMA-AM 420
 QY 371 EK--DQLAQLEE-KDLL-----QKQVNEKIEENLTRMLV- 402
 DB 421 QKTDDEKKALEEKDMVEEDRNTVHRELQRSESLHKAQDDQKILNEKLAIOKILVG 480
 QY 403 -----SSSLTLOELKAKRKRVTVCLGKINKMKNYADQFNPTNTTK 446
 DB 481 GVDLLAKSEQEQLEQSALEKMERMAKQESMR-----KMMEEERQERMDIEEKYS 531
 QY 447 T-----KTKL 452

[illegible]

Db 487 ASHMKTEQVLTTQACEILAAADLATDTHQLHGTIERRRELDEKTRRSQDFKRMQDNL 546

QY 479 SEI 481

Db 547 EMI 549

RESULT 14

H86350

hypothetical protein F8K7.17 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001

C:Accession: H86350

R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Jensen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: H86350

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-909 <STO>

A:Cross-references: GB:AE005172; NID:g5263326; PIDN:AAD41428.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

Query Match 28.2%; Score 700; DB 2; Length 909;

Best Local Similarity 39.3%; Pred. No. 1.2e-32;

Matches 169; Conservative 86; Mismatches 143; Indels 32; Gaps 10;

QY 6 VAVCVVRPLNRSBSLGETAQVYWKTDNNVIVQDGSKSFNDRVFHGNETTKNVYEEI 65

Db 76 ITVTIRFRLPRLPVNNGDEIAWADGYTIRNEYNPSLCVGFDRVFGPTTTRRVVDIA 135

QY 66 AAPIDSAIOGYNGTTFAYGQTASGKTYTWMSGDHLGVTIPRAIHDIFQIKKF----- 119

Db 136 AQOVSGAMSGINGTTFAYGVTSKGTHHNGQRSGPIPLAVKDVFSIIQEVIANVYS 195

QY 120 -----PDRFLLRYSYMEIYNETITDLCGTQRMKPLIIRRDVARNVYVADLTVEEYVTS 175

Db 196 AIOPTEREFLLRVSYLEIYNEVINDLDPGT--QNLRIED-SQGIYVEGIGDEVLSPA 252

QY 176 MALKWLTKE---KSRHYGETKQNRSSRSHITFRMILSRKGEFSPNCGSVKVSHLNL 232

Db 253 HALSLTASGEVIAEHRHVGSNNVLFSSRSHMTFTLTIESPHRGKGDGE-DVLSQLHL 311

QY 233 VDLAGSRAAQTAGAAGVRLKEGNCINRSILFILGOVTKLSDGVGGFINVRDSKLTRLQ 292

Db 312 IDLAGE-SSEKTEITQRRKEGSSINKSLLTGLTVLSKLTDTK-AAHIPYRDSKLTRLQ 369

QY 293 NSLGNPKTRITCTTPVS--FDELTALQFSTAKYMKNTFYVNEVSTDEALLKRYKE 350

Db 370 STLSGHGRVSLICTTPASTSEETHNLKFAQRCKHVEIKASRNKIMDEKSLIKKYQKE 429

QY 351 IMDLKQQLKEV-----SLETRAAMEKDQLAQL-----LEEKLLQKQVNEKLENLTRM 399

Db 430 ISCLQELTQLRHGNQDLDLRKLQVGKLSQVKKLQSLQSLQSLQSLQSLQSLQSLQSLQ 489

QY 400 LVTSSSLTLQ 409

Db 490 ILVSTKSSLQ 499

RESULT 15
T21134
hypothetical protein F20C5.2 - *Caenorhabditis elegans*
C;Species: *Caenorhabditis elegans*


```

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T21134
R:Matthews, P.
submitted to the EMBL Data Library, December 1995
A:Reference number: Z19381
A:Accession: T21134
A>Status: Preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-1130 <WIL>
A:Cross-references: EMBL:Z68161; PIDN:CAA92295.1; GSPDB:GN00022; CESP:F20C5.2
A:Experimental source: clone F20C5
C:Genetics:
A:Gene: CESP:F20C5.2
A:Map position: 4
A:Introns: 15/3; 34/2; 69/2; 152/3; 183/3; 207/2; 314/3; 513/3; 546/2; 594/1; 632/3; 667/2

Query Match      28.2%; Score 699; DB 2; Length 1130;
Best Local Similarity 34.8%; Pred. No. 1.8e-32;
Matches 194; Conservative 87; Mismatches 174; Indels 102; Gaps 16;

Qy 1 AEGAVAVCVVRPLNSRESLSGETAQTAVYKTDNNVI-----YQVDSKSFNDFRVFHG 54
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 55 NETTKNVEETAAPIIDLSAQYNGTIPAYGQTASGKTYTMGSE---DHLGVIPRAIHD 111
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 69 NSTQSDLYEETFRDLVDSVLNGYNATIFAYGQTGTGKTHMEGKSSDPEQRGVIYKCIDH 128
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 112 IFQIKKFPDPREFLLRVSYMEIYNETITDLGCGTKMKPLIREDVNRNVTADLTREV 171
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 129 IFEHMAASHNQYLVASYLEIYQELRDLL-EAESNKKLEIKERPDPGGVYVKOLTSKLT 187
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 172 YTSEWALKWITKGKSRHYGTQNMORSRSHITPFRMILESRKGEPSNCEGSKVSHLN 231
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 188 RTVGEIHEVMIRGNGHRSVGRVTNMEHSRSHAFIITVECSRIGEDG--ESHITVGRLN 245
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 232 LVDLAGSRAACTGAAGVRLKEGCNINRSLFILGOVKKLSGQVGGFINYRDSKLTREL 291
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 246 LVDLAGSERQSKTGATGERFEATKINLSALGNVISALVDK-SAHIPYRDSKLTREL 304
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 292 QNSLGGNPKTRICTITPVSFDELTALQFSTAKYKNTPYVNEVSTDEALLKRYKEL 351
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 305 QDSLGGNSKT-----ETLGLTRYANRAKNIKQPKINBDPKD-ALLREFQEEI 351
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 352 MDLKKQL-----EVSLETRAQAM-----EKDQLAOLLEEK-D 383
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 352 EMLRQLKQRKTRSDGATQSFYDAERAKLEDIEAIQKDDSLIKHEKDLIREIQEKHD 411
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 384 LIQKVQ-----NEKIENTRLMTSS-----SLTLQOELKAKRKRRVTWCLGKINNM 430
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 412 LLEKERIEQARVAERIANIQSLIVGSEEDGRLESRTKEQHAQLEKKRR-----ELAEQ 465
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 431 K-----NSNVADQNPINPTNITTKTKLSINLLR-----BIDESV 464
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 466 KRREREMVEALERQEBDVTDLQKTFSD---LRTVEAKTKKLKQMLIKLRQARNEIRDVS 522
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 465 CSESDFVFNLTDLTSLSEI 481
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 523 GAYSDEKDLDDQTIAEV 539
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Search completed: July 29, 2004, 09:41:06
Job time : 8.27641 secs

Result No.	Query No.	Query			DB	ID	Description
		Score	Match	Length			
1	1	2483	100.0	2663	1	CENE HUMAN	Q02224 homo sapien
2	2	736.5	29.7	747	1	KE3B_HUMAN	O15066 homo sapien
3	3	735.5	29.6	747	1	KE3B_MOUSE	Q61771 mus musculu
4	4	728	29.3	701	1	KE3A_HUMAN	Q28741 mus musculu
5	5	723	29.1	702	1	KE3A_MOUSE	Q9Y496 homo sapien
6	7	722	29.1	742	1	K121_STRPU	P46871 strongyloce
7	7	716	28.8	786	1	FL10_CHIRLE	P46869 chlamydomon
8	7	713.5	28.7	1232	1	KE4A_HUMAN	P59239 homo sapien
9	9	710.5	28.6	1029	1	KF17_HUMAN	Q9P2e2 homo sapien
10	11	710.5	28.6	1231	1	KP4A_MOUSE	P33174 mus musculu
11	10	704.5	28.4	699	1	K122_STRPU	P46872 strongyloce
12	12	703	28.3	1066	1	KL61_DROME	P46863 drosophila
13	13	702.5	28.3	1038	1	KE17_MOUSE	Q99pw8 mus musculu
14	14	692	27.9	1226	1	KF4A_XENLA	Q91784 xenopus lae
15	15	676.5	27.2	1225	1	KP4A_CHICK	Q30640 gallus gall
16	16	664.5	26.8	1031	1	KINH_STRPU	P35978 strongyloce
17	17	660.5	26.6	975	1	KINH_DROME	P71210 drosophila
18	18	656	26.4	672	1	OSM3_CAEEL	P46873 caenorhabdi
19	19	656	26.4	796	1	KE3C_RAT	O55165 rattus norv
20	20	655	26.4	796	1	KE3C_MOUSE	Q35066 mus musculu
21	21	654.5	26.4	963	1	KINH_HUMAN	P33176 homo sapien
22	22	653.5	26.3	963	1	KINH_MOUSE	Q61768 mus musculu
23	23	652.5	26.3	1749	1	K13A_MOUSE	Q9eqw7 mus musculu
24	24	652	26.3	793	1	KE3C_HUMAN	O14782 homo sapien
25	25	652	26.3	1805	1	K13A_HUMAN	Q9h1h9 homo sapien
26	26	650	26.2	1826	1	K13B_HUMAN	Q9nqt8 homo sapien
27	27	642.5	25.9	967	1	KINH_LOLPE	P21613 loligo peal
28	28	640	25.8	957	1	KP5C_HUMAN	O60282 homo sapien
29	29	638.5	25.7	1032	1	KINN_HUMAN	Q12840 homo sapien
30	30	636.5	25.6	956	1	KE5C_MOUSE	P28738 mus musculu
31	31	635.5	25.6	1027	1	KINN_MOUSE	P33175 mus musculu
32	32	633	25.5	784	1	KL68_DROME	P46867 drosophila
33	33	632.5	25.5	1057	1	KF11_HUMAN	P52732 homo sapien

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; Z15005; CAA78727.1; -

PIR; S28261; S28261.

HSSP; P17119; 3KAP.

Genew; HGNC:1856; CENPE.

GK; Q02224; -

MIM; 117143; -

GO; GO:0005699; C:kinetochore; TAS.

GO; GO:0005634; C:nucleus; TAS.

GO; GO:0008350; P:kinetochore motor activity; TAS.

GO; GO:0000067; P:DNA replication and chromosome cycle; TAS.

GO; GO:0007079; P:mitotic chromosome movement; TAS.

GO; GO:0007080; P:mitotic metaphase plate congression; TAS.

InterPro; IPR001752; kinesin_motor.

Pfam; PF00225; kinesin; 1.

PRINTS; PR00380; KINESINHEAVY.

SMART; SM00129; KISC; 1.

PROSITE; PS00411; KINESIN_MOTOR_DOMAIN; 1.

PROSITE; PS00667; KINESIN_MOTOR_DOMAIN; 1.

Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;

Cell cycle; Centromere; Lipoprotein; Prenylation.

DOMAIN 1 335 KINESIN-MOTOR.

DOMAIN 336 2471 COILED COIL (POTENTIAL).

DOMAIN 2472 2663 GLOBULAR (POTENTIAL).

NP_BIND 86 93 ATP (BY SIMILARITY).

LIPID 2660 2660 S-farnesyl cysteine.

SEQUENCE 2663 AA; 312087 MW; CEFCL3880C8C8B8 CRC64;

Query Match 100.0%; Score 2483; DB 1; Length 2663;

Best Local Similarity 100.0%; Pred. No. 3.1e-136;

Matches 487; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEEGAVAVCVVRPLNSREESLGETAQVYKNTDNNVYQVDGSKSFNDRFVHGNETTKN 60
 Db 2 AEEGAVAVCVVRPLNSREESLGETAQVYKNTDNNVYQVDGSKSFNDRFVHGNETTKN 61
 QY 61 VYEEIAAPIIDSAIQGYNGTTFAYGQTASGKTYTMGSEDLGVIPRAIHDFQKIKFF 120
 Db 62 VYEEIAAPIIDSAIQGYNGTTFAYGQTASGKTYTMGSEDLGVIPRAIHDFQKIKFF 121
 QY 121 DRELLRVSYMEIYNETITDLCGQKMKPLIREDVNRNYYVADLTVEEVYTTSEMALKW 180
 Db 122 DRELLRVSYMEIYNETITDLCGQKMKPLIREDVNRNYYVADLTVEEVYTTSEMALKW 181
 QY 181 ITKGEKSRHYGETKXNQSSRSHTIFRMILSRKGEPSNCEGSKVYSHLNLVDLAGSER 240
 Db 182 ITKGEKSRHYGETKXNQSSRSHTIFRMILSRKGEPSNCEGSKVYSHLNLVDLAGSER 241
 QY 241 AAOQTGAAGVRUKEGCNINRSFILGQVTKKLSGQGVGFNRYRSKLTIRLONSLGGNPK 300
 Db 242 AAOQTGAAGVRUKEGCNINRSFILGQVTKKLSGQGVGFNRYRSKLTIRLONSLGGNPK 301
 QY 301 TRILCTITPVSFDETLTALQPASTAKYKNTPYNEVSTDEALLKRYRKEIMDLKKOLEE 360
 Db 302 TRILCTITPVSFDETLTALQPASTAKYKNTPYNEVSTDEALLKRYRKEIMDLKKOLEE 361
 QY 361 VSLETRAAMEKDQALAEKDLQKQNEKINLTMLVTSSSLTIQBELKAKRKRRV 420
 Db 362 VSLETRAAMEKDQALAEKDLQKQNEKINLTMLVTSSSLTIQBELKAKRKRRV 421
 QY 421 TWCLGKINKMKNVADQFNIPNTITTKLISNLLREIDSVCSDFVNTLDTLSE 480
 Db 422 TWCLGKINKMKNVADQFNIPNTITTKLISNLLREIDSVCSDFVNTLDTLSE 481
 QY 481 LEWNPAT 487
 Db 482 LEWNPAT 488

RESULT 2

KF3B HUMAN

ID KF3B HUMAN

STANDARD; PRT; 747 AA.

AC Q15066;

DT 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Kinesin-like protein KIF3B (Microtubule plus end-directed kinesin

motor 3B) (HH0048).

DE KIF3B OR KIAA0359.

GN KIF3B OR KIAA0359.

OS Homo sapiens (Human)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.

OX NCBI_TaxID=9606;

[1]

SEQUENCE FROM N.A.

PC TISSUE=Brain;

PC MEDLINE=97349984; PubMed=9205841;

RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,

RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;

RT "Prediction of the coding sequences of unidentified human genes. VII.

RT The complete sequences of 100 new cDNA clones from brain which can

RT code for large proteins in vitro.";

RL DNA Res. 4:141-150(1997).

[2]

SEQUENCE FROM N.A.

RP MEDLINE=21638749; PubMed=11780052;

RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,

RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baguley C.L.,

RA Bailey J., Barlow K.F., Bates K.N., Bead L.M., Beare D.M.,

RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,

RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P., Clee C.M.,

RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,

RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,

RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,

RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,

RA Grahnam D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,

RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,

RA Huckle E., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,

RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,

RA Lehaeslaio M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,

RA Marsh V.L., Martin S.L., McConnachie L.J., McEay K., McMurray A.A.,

RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,

RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,

RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,

RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkeen R., Sims S.,

RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,

RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,

RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,

RA Whitehead S.B., Whittaker P., Willey D.L., Williams L., Williams S.A.,

RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,

RA Rogers J.;

"The DNA sequence and comparative analysis of human chromosome 20.";

Nature 414:865-871(2001).

[3]

IDENTIFICATION IN A COMPLEX WITH SMC3 AND KIFAP3B.

MEDLINE=98175913; PubMed=9506951;

RA Shimizu K., Shiratani H., Honda T., Minami S., Takai Y.;

RA "Complex formation of SMAP/KAP3, a KIF3A/B ATPase motor-associated

protein, with a human chromosome-associated polypeptide.";

J. Biol. Chem. 273:6591-6594(1998).

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

SEQUENCE FROM N.A.
 TISSUE=Testis;
 MEDLINE=2388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schaefer C.F., Bhat N.K.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McWay P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 [3]
 SEQUENCE FROM N.A.
 RA Connolly K.S., Gunning K.M., Davis C.A., Kadner K., Subramanian S.,
 RA Miguel T., Lewis K.D., Fridlyand J., Alcivare D., Benke J.A.,
 RA Bondoc M., Bowen E., Chiang A., Critz P., Jaklevic M.A., Lindo K.,
 RA Lindquist K., Miller C., Patel S., Plascia C., Riley B.E., Rojeski H.,
 RA Sarmiento R., Yu C., Montenegro M., Aerts A., Chung A., Abramo A.,
 RA Baker M., Gau C., Jett J., Ko C., Beall K., Woolley J.P., Stultz J.L.,
 RA Kimmerly W., Martin C.H.;
 RT "Sequencing of human chromosome 5";
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: MICROTUBULE-BASED ANTEROGRADE TRANSLATOR FOR
 MEMBRANOUS ORGANELLES. PLUS END-DIRECTED MICROTUBULE SLIDING
 ACTIVITY IN VITRO.
 CC -!- SUBUNIT: HETERODIMER OF KIF3A AND KIF3B.
 CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
 II SUBFAMILY.
 CC -!- CAUTION: Ref.3 sequence differs from that shown due to erroneous
 gene model prediction.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 DR EMBL; AF041853; AAC72294.1; --
 DR EMBL; BC045542; AAH45542.1; --
 DR EMBL; AC004237; AAC04475.1; ALT_SEQ.
 DR HSSP; P17119; 3KAR.
 DR Genew; HGNC:6319; KIF3A.
 DR MIM; 604683; --
 DR GO; GO:0006996; P:organelle organization and biogenesis; TAS.
 DR InterPro; IPR001752; kinesin_motor.
 DR Pfam; PF00225; kinesin; 1.
 DR PRINTS; PR00380; KINESINHEAVY.
 DR SMART; SM00129; KISC; 1.
 DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
 DR PROSITE; PS00067; KINESIN MOTOR DOMAIN2; 1.
 KW Motor protein; Microtubule; ATP-binding; Coiled coil; Neurone.
 FT DOMAIN 1 350 KINESIN-MOTOR.
 FT DOMAIN 351 599 COILED COIL (BY SIMILARITY).
 FT DOMAIN 600 702 GLOBULAR.
 FT NP_BIND 100 107 ATP (BY SIMILARITY).
 FT DOMAIN 443 446 POLY-GLU.
 FT CONFLICT 151 151 E -> G (IN REF. 2).
 FT CONFLICT 170 170 E -> A (IN REF. 1).
 FT CONFLICT 172 172 K -> I (IN REF. 2).

SQ SEQUENCE 702 AA; 80385 MW; 88555D21209B4E14 CRC64;
 Query Match 29.1%; Score 723; DB 1; Length 702;
 Best Local Similarity 35.8%; Pred. No. 8.5e-35;
 Matches 193; Conservative 84; Mismatches 166; Indels 96; Gaps 14;
 QY 6 VAVCVVRPLNSRSL--GETAQVYVKKTDNNVIVQVDS-----KSFNDFRPHGNETTK 59
 DB 15 VKVVVRCPPLNERKSCVQKQVSVDEMGTTVHKTDSSNEPPKTFTEFTVFGPESKOL 74
 QY 60 NYVEEIAPIIDSAIQNGNTIFAYGQTASGTYTMMGSE---DHILGVIPRAIHDFQKI 116
 DB 75 DYNLTARPIIDSVLENGNTIFAYGQTGKTFTMEGVRAIPELRGITPNSFAHIFGHI 134
 QY 117 KFPF--DRFELLVSVMEYNETITDLCQTQOMKPLIIREDVNRNMYVADLFEVVYTSSE 175
 DB 135 AKAEGETRFLVRVSVLEIYNEEVRDL--CKDQTORLEVKERPDVGVIYIKLSAYVNNAD 193
 QY 176 MALKWITKGEKSRHYGETKMNORSSRSHTIPRMILESEKGPSPCEGSKVYSHLNLVDL 235
 DB 194 DMDRIMTILGHKNSRVGATNMHSHSRSHAFITITIECKEKIDGNWH--VRMGKLUHLVDL 251
 QY 236 AGSEAAQTGAAGVRLKEGCNINRSLFIIGQVKKLSGQGVGFNYRDSKILTRILQNSL 295
 DB 252 AGSERQAKTGATGQRLKEATKINLSLSTLGNVISALVDGK--STHVPYENSKLTRILQDSL 310
 QY 296 GGNPKTRIICTITPV--SFDETLTALQASTAKYMNTPYNEVSTDEALLKRYRKEIMD 353
 DB 311 GGNSTWMCANIGPADYNYDETISTLYRANAKNIKRNARINEDPKD--ALLRFQKEIEE 369
 QY 354 LKKQLEB----- 360
 DB 370 LKKLEEGEEISGDISSEDDDEGEVGEKRRKRRDQTGKKKVSPPDMIMQAKI 429
 QY 361 ----VSLETRAQAMEKDLQALLE----EKOLQKQVNEKIEINLRMLVTSSSLTLOQEL 412
 DB 430 DEERKALETKLDMEERNEERKARAELEKREKOLLKAQAEHQ-----SLEKL 475
 QY 413 KAKRRRVTWCLGINKMKNYADQFNIPNI--TTTKHKLSINLLRIDSVCSSESDV 470
 DB 476 SALEKKVI---VGGVDLLAKAEQEKLEESNMELERRKRAEQLRRELEKQERLDOI 531
 RESULT 6
 KI21_STRPU STANDARD; PRT; 742 AA.
 ID AC P46871;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Kinesin-II 95 kDa subunit (KRP-85/95 95 kDa subunit).
 GN KRP95.
 OS Strongylocentrotus purpuratus (Purple sea urchin).
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
 OC Echinoidea; Euechinoidea; Echinacea; Echinoidea; Strongylocentrotidae;
 OC Strongylocentrotus.
 OX NCBI_TaxID=7668;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Egg;
 RX MEDLINE=94050179; PubMed=8232586;
 RA Cole D.G., Chinn S.W., Wedaman K.P., Hall K., Vuong T.,
 RA Scholey J.M.;
 RT "Novel heterotrimeric kinesin-related protein purified from sea
 urchin eggs";
 RL Nature 366:268-270 (1993).
 CC -!- SUBUNIT: Heterotrimer of a 115 kDa subunit (KAP115) and two
 CC kinesin-like subunits of 95 kDa (KRP95) and 85 kDa (KRP85).
 CC -!- PTM: The N-terminus is blocked.
 CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
 II SUBFAMILY.

 This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; U00996; AAA87393.1; --
CC HSPSP; P17119; 3KAR.
CC InterPro; IPR001752; kinesin_motor.
CC Pfam; PF00225; kinesin; 1.
CC PRINTS; PR00380; KINESINHEAVY.
CC SMART; SM00129; KISC; 1.
CC PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
CC PROSITE; PS0067; KINESIN MOTOR DOMAIN2; 1.
CC Motor protein; Microtubule; ATP-binding; Coiled coil.
CC DOMAIN 1 337 KINESIN-MOTOR (BY SIMILARITY).
CC FT DOMAIN 338 COILED COIL (BY SIMILARITY).
CC FT DOMAIN 613 COILED COIL (BY SIMILARITY).
CC FT NP BIND 95 102 GLOBULAR (BY SIMILARITY).
CC FT NP BIND 95 102 ATP (POTENTIAL).
CC SEQUENCE 742 AA; 84202 MW; 47C40A367BAA77B5 CRC64;

Query Match 29.1%; Score 722; DB 1; Length 742;
Best Local Similarity 37.6%; Pred. No. 1e-34;
Matches 198; Conservative 74; Mismatches 191; Indels 64; Gaps 13;

QY 6 VAVCVVRPLNSRESLGESLGTAYQVYKTDNNVIYQV-----DGSKSPNDRVPHGNETT 58
DB 9 VKVYVRCRPLNGKEKADGRSRIYDMDVDAGQVVRPNPKADASEPPKAFTFDQVYDMWCQ 68
QY 59 KNYVEEIAAPIDSAGIQVNGTIFAYGQTASGKTYTMWG---SEDLGLVPIRAIHDFQK 115
DB 69 IDLYDETRSLVESVLQGFNGTIFAYGQTGSGKTYTMWG---SEDLGLVPIRAIHDFQK 128
QY 116 IKKPPDREFLLRVSYMEIYNETITDLCCTQKPKLIIREDVNRNVYVADLVEEVYV 175
DB 129 IARTNQOQFLVRASYLEIYQEIREDLLAKQK-KRLDLKERPDGTGVYKOLSSFTVSVK 187
QY 176 MALXWITKGEKSRHYGETKMNQSRSHITFRMILESRKGEPCNSGVKVSHLNLDL 235
DB 188 EIEHVMVTGNNRSVSGTNNMHSRSHAFITIECSLGVGK--ENHIVGKLNLDL 245
QY 236 AGSERRAQTGAAGVRLKEGCNINRSLFILGQVVKLSGQVGGFINVRDSKILRLQNSL 295
DB 246 AGSERQATGATGDLKATKINLSLALGNSALVSDGK--SHIYEDSKILRLQNSL 304
QY 296 GGNPKTRICTITPVV---FDETLALQFASAKYKNTPTPYNEVSTDBALLKRYKKEIMD 353
DB 305 GGNKATVMVANMGFASYNFDETTITLRYANRAKINKPKINEDPKD--ALLREFFQEISR 363
QY 354 LKKQL-----EEVSLTRAQAMEKQDLAOLLEKD 383
DB 364 LKQALDKKPGSDGRKKKRPGEQGGDDIEDTEEGEDMEDEEMVYKESQKLEBEKE 423
QY 384 LLQVQNEKIEMLTMLVTSSLTILQBELKAKRRVTVICGLKINKMKNKNYAQFNPT 443
DB 424 KIMANQSMABEKQKL--SEVQKQGEIK-KEHQCKEMLEKGIKAMESKLLVGGKSIDV 480
QY 444 NITTKHLSINLL-----REID-ESVCSSESVFNTDLTSLIEIR 482
DB 481 HTNEQQRXIEQRLLABEKNRERDMERLKEQD-----DKTVIEIE 521

RESULT 7

ID FL10_CHLRE STANDARD; PRT; 786 AA.
AC P46869;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Kinesin-like protein FLA10 (KIF1 protein).
GN FLA10.
OS Chlamydomonas reinhardtii.

OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=J37;
RX MEDLINE=94299638; PubMed=8027176;
RA Walther Z., Vaehishtha M., Hall J.L.;
RT "The Chlamydomonas FLA10 gene encodes a novel kinesin-homologous
protein";
RL J. Cell Biol. 126:175-188(1994).
CC -!- FUNCTION: Probably involved in flagellar assembly and maintenance.
CC May play a role in flagellar synthesis.
CC -!- TISSUE SPECIFICITY: Flagellar axoneme.
CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
CC II SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

EMBL; L33697; AAA21738.1; --
PIR; A53939; A53939.
HSPSP; P17119; 3KAR.
InterPro; IPR001752; kinesin_motor.
Pfam; PF00225; kinesin; 1.
PRINTS; PR00380; KINESINHEAVY.
SMART; SM00129; KISC; 1.
PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
PROSITE; PS0067; KINESIN MOTOR DOMAIN2; 1.
Motor protein; Microtubule; ATP-binding; Coiled coil.
DOMAIN 1 358 KINESIN-MOTOR (BY SIMILARITY).
DOMAIN 367 687 COILED COIL (POTENTIAL).
DOMAIN 688 786 GLOBULAR (POTENTIAL).
NP BIND 97 104 ATP (POTENTIAL).
DOMAIN 388 391 POLY-GLY.
DOMAIN 705 714 POLY-GLY.
DOMAIN 756 759 POLY-ASP.
SEQUENCE 786 AA; 86671 MW; F90969203EB79F1B CRC64;

Query Match 28.8%; Score 716; DB 1; Length 786;
Best Local Similarity 36.3%; Pred. No. 2.5e-34;
Matches 194; Conservative 86; Mismatches 176; Indels 78; Gaps 16;

QY 5 AVAVCVVRPLNSRESLGESLGTAYQVYKTDNNVIY-----QVDGS---KSFNDRVPHGNET 57
DB 10 SVKVVVRCRPLNGKEKADGRSRIYDMDVDAGQVVRPNPKADASEPPKAFTFDQVYDMWCQ 69
QY 58 KNYVEEIAAPIDSAGIQVNGTIFAYGQTASGKTYTMWGSEHDH---LGVIPIRAIHDFQ 114
DB 70 QRDVFDITARLIDSCIEGNGTIFAYGQTGSKSHMEGKDEPELRLGLIPNTFRYVFE 129
QY 115 KI-KKFPDREFLLRVSYMEIYNETITDLCCTQKPKLIIREDVNRNVYVADLVEEVYV 173
DB 130 IIRDSGTKEFLVRSSYLEIYNEVROLL-GKDSKMKMELKESPDGVYVVDLSQFVCKN 188
QY 174 SEMALKWITKGEKSRHYGETKMNQSRSHITFRMILESRK-----GEPSNCEGS 224
DB 189 YEENKVKLLAGKDRNRQVATLMNQDSRSRSHIFITIECIKLSAAAKQKGAKKDDSNH 248
QY 225 VKVSHLNLVDLAGSERRAQTGAAGVRLKEGCNINRSLFILGQVVKLSGQVGGFINVRD 284
DB 249 VRVGKLNLDLAGSERQDKTGTGDLKIGIKINLSLALGNSALVSDGK--SGHIPYRD 307
QY 285 SKLFRILQNSLGGNPKTRICTITPVV---SFDETLALQFASAKYKNTPTPYNEVSTDBEA 342
DB 308 SKLFRLLQDSLGSNTKTYVNAVIGVADNWDYETMTSLRYANRAKINKPKINEDPKD-A 366
QY 343 LLKRYKKEIMDLKKQL-----EEVSLTRAQAMEKQD 374

Db 367 MLRQFOEIKKLEQAARAAGGGPITMPSGGSPQTKIVERTTEVDPIDA---IKAQ 423
 QY 375 LAQLLEK-----DLQKVQNE---KLENLRLMTVSSSLTLOQLKA-----KKRR 419
 Db 424 MRLEAKMKSDISTALDKAREEAAAKKQLQAIIDQCKTEAOKKAARDALKKQAAE 483
 QY 420 VTWCLGKINKMKNVADQFNPIITTKHKU---SINLLREIDSVCSSESV 470
 Db 484 ARAIAGAIEKEQ-----EKAVLESRIKMEGKIVGGVNMLEKVDLQKQSEDI 533

RESULT 8

KF4A_HUMAN
 ID KF4A_HUMAN STANDARD; PRT; 1232 AA.
 AC Q95239; Q9NNY6; Q9NY24; Q9UMW3;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 16-OCT-2001 (Rel. 40, last sequence update)
 DT 28-FEB-2003 (Rel. 41, last annotation update)
 DE Chromosome-associated kinesin KIF4A (Chromokinesin).
 GN KIF4A OR KIF4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lymphocytes;
 RA Villard L.;
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
 RX MEDLINE=20435301; PubMed=10978527;
 RA Oh S.J., Hahn H., Torrey T.A., Shin H., Choi W., Lee Y.M.,
 RA Worsle H.C. III, Kim W.;
 RT "Identification of the human homologue of mouse KIF4, a kinesin
 superfamily motor protein.";
 RL Biochim. Biophys. Acta 1493:219-224(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Retinoblastoma;
 RA Rentsch A., Neumann T., Rommerskirch W.;
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 128-1232 FROM N.A.
 RC TISSUE=Retinoblastoma;
 RX MEDLINE=97911419; PubMed=9168136;
 RA Yan R.-T., Wang S.-Z.;
 RT "Increased chromokinesin immunoreactivity in retinoblastoma cells.";
 RL Gene 189:263-267(1997).
 CC -!- FUNCTION: REQUIRED FOR MITOTIC CHROMOSOMAL POSITIONING AND BIPOLAR
 CC SPINDLE STABILIZATION (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: Nuclear. Associated with mitotic
 CC chromosomes (by similarity).
 CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN HEMATOPOIETIC TISSUES,
 CC FETAL LIVER, SPLEEN, THYMUS AND ADULT THYMUS AND BONE MARROW.
 CC LOWER LEVELS ARE FOUND IN HEART, TESTIS, KIDNEY, COLON AND LUNG.
 CC -!- SIMILARITY: Belongs to the kinesin-like protein family.
 CC Chromokinesin subfamily.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sb.ch).

 DR EMBL; AF179308; AAD51855.1; -;
 DR EMBL; AF071592; AAD05492.2; -;
 DR EMBL; AJ271784; CAB75427.1; -;
 DR EMBL; AF277375; AAF86334.1; -;
 DR HSSP; P17119; 3KAR.

Genew; HGNC:13339; KIF4A.
 GO; GO:0005737; Cytoplasm; TAS.
 GO; GO:0005876; C-spindle microtubule; TAS.
 GO; GO:0003777; F-microtubule motor activity; TAS.
 GO; GO:0008089; P-antegrade axon cargo transport; TAS.
 GO; GO:0006996; P-organelle organization and biogenesis; TAS.
 InterPro; IPR001752; kinesin_motor.
 Pfam; PF00225; kinesin; 1.
 PRINTS; PRO0380; KINESINHEAVY.
 SMART; SM00129; KISC; 1.
 DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
 DR PROSITE; PS0067; KINESIN MOTOR DOMAIN2; 1.
 KW Motor protein; Microtubule; ATP-binding; DNA-binding;
 KW Nuclear protein; Coiled coil.
 FT DOMAIN 1 349 KINESIN-MOTOR.
 FT DOMAIN 350 999 COILED COIL (BY SIMILARITY).
 FT DOMAIN 1000 1232 GLOBULAR.
 FT NP_BIND 88 95 ATP (POTENTIAL).
 FT CONFLICT 223 223 R -> G (IN REF. 2).
 FT CONFLICT 231 231 S -> T (IN REF. 4).
 FT CONFLICT 286 286 V -> A (IN REF. 2).
 FT CONFLICT 422 422 L -> W (IN REF. 2).
 FT CONFLICT 564 564 L -> H (IN REF. 4).
 FT CONFLICT 564 564 L -> P (IN REF. 2).
 FT CONFLICT 600 600 K -> E (IN REF. 3).
 FT CONFLICT 668 668 R -> K (IN REF. 3 AND 4).
 FT CONFLICT 928 928 Q -> P (IN REF. 1).
 FT CONFLICT 958 958 Q -> R (IN REF. 3).
 FT CONFLICT 960 960 L -> Q (IN REF. 1).
 FT CONFLICT 996 997 LL -> S (IN REF. 4).
 FT CONFLICT 1003 1014 OKHLPKDTLLSP -> RYLPRIPFFYLQ (IN REF. 4).
 FT CONFLICT 1022 1022 P -> Q (IN REF. 2).
 FT CONFLICT 1077 1077 K -> N (IN REF. 2).
 FT CONFLICT 1138 1138 G -> S (IN REF. 2).
 SQ SEQUENCE 1232 AA; 139908 MW; FF74052A17A8E8F7 CRC64;
 Query Match 28.7%; Score 713.5; DB 1; Length 1232;
 Best Local Similarity 35.1%; Pred. No. 6.3e-34;
 Matches 182; Conservative 86; Mismatches 191; Indels 59; Gaps 11;
 QY 6 VAVCVVRPLNSRRESLG-ETAAQVYWKTDNNVIYQVDSKSFNDFRVPFHGNETTKNVEE 64
 Db 10 VRVALRCPLVPEKEISEGCMCLSFVPGEQVVGTD--KSTYDFVFPSTEQEVEFT 67
 QY 65 TAAPIDSIAIQNGTIFAYGQTASGKYTYMMG-----SEHLGVIPRAIHDFQKTK 117
 Db 68 AVAPLKGKGVNATVLAIGTSGKTYSMGAYTAEOENEPYGVIPRVQLLFKEID 127
 QY 118 KFPDREFLLRVSMEIYNETTDLCCGQKPKLIREDVNVNRYVADLTVEWVYSEMA 177
 Db 128 KKSDFEFTLKVSYLEIYNEELDLCCPSREKAQINIREDPKEGIVGLTEKTVLVALDT 187
 QY 178 LKWITGKESRHYGTEKNQSRSSHTIFRMILESEKGEPSNCEGSKVSHNLVLDLAG 237
 Db 188 VSCLEQGNNSRTVASTAMNSQSSRSHAFITISLEQRKSD---KNSFSRKLHLVDLAG 243
 QY 238 SERAAQTAAGVRLKEGCNINRSLPILQVTKLSDGVGVGFYINRDSKLTILQNSLGG 297
 Db 244 SERQKTKAEGDRLKEGINIRGLICLGNVISALGDDKGGFVPYRDSKLTLLQDSLGG 303
 QY 298 NPKRIICTIPV--SPDETALQFASTAKYMKNTPVNVESTDALLKRYRKIMDLK 355
 Db 304 NSHTLMIAVCSPADSNLEETINTURYADRARKINKKPIVN-----IDPOTAEHLNHLK 355
 QY 356 KQLREVS!-----ETRAQAMEKDLAQLLEKDLLOKQVONKIEML 396
 Db 356 QVQQLQVLLLAQHGGTLPGSITVEPSENLQSLMEKNQ--SLIVENEKLRLGLSFAAQGT 413
 QY 397 TRML---VTSSSLTLQOELKAKRKRVTCWLGKINK-----MKNSNYADQFNPTNITTKT 449
 Db 414 AQMLERIILTQANEKKNNAKLEELRQHAACKLDLQKLVTETLEDQELKENVEIICNLQOLI 473
 QY 450 HKLSINLLREIDSVCSSESVFNSLTDLTSLSEIWNPAT 487

QY 349 KEIMDLKKOLEVSLLETRAQAMEKDKQAQLL-----BEKOLLQKV----- 388
 Db 355 ERIKLLKAIL-----TQMGSPSLLSALLSQVPPDPVQVEKLLPQVIGHDMEAEK 406
 QY 389 ---ONEKIENTRLMVLTSLSLITLOELKAKRKRVVWCLGKINKNKNSVADOFNPTNI 445
 Db 407 OLIREYEBERLRL---KADYKAEQSRARLEEDIT-----AMRNSYDVRSLTLEENL 456
 QY 446 TTXTKLSINLLREIDESVCSDFNSLTDLTLSLETWNP 486
 Db 457 RKETEAV-----LQGVLYKAEVMSRA-EFASSAEYPPA 489

RESULT 10
 KF4A MOUSE
 ID KF4A MOUSE STANDARD; PRT; 1231 AA.
 AC P33174;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Chromosome-associated kinesin KIF4A (Chromokinesin).
 GN KIF4A OR KIF4 OR KNS4.
 OS Mus musculus (Mouse)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, AND TISSUE
 RP SPECIFICITY
 RC STRAIN=ICR; TISSUE=Brain;
 RX MEDLINE=95014709; PubMed=7929562;
 RA Sekine Y., Okada Y., Noda Y., Kondo S., Aizawa H., Takemura R.,
 RA Hirokawa N.;
 RT "A novel microtubule-based motor protein (KIF4) for organelle
 RT transports, whose expression is regulated developmentally.";
 RL J. Cell Biol. 127:187-201(1994).
 RN [2]
 RP SEQUENCE OF 91-240 FROM N.A.
 RC TISSUE=Brain;
 RA Aizawa H., Sekine Y., Takemura R., Zhang Z., Nangaku M.,
 RA Hirokawa N.;
 RT Kinesin family in murine central nervous system.";
 RL J. Cell Biol. 119:1287-1296(1992).
 CC -!- FUNCTION: REQUIRED FOR MITOTIC CHROMOSOMAL POSITIONING AND BIPOlar
 CC SPINDLE STABILIZATION.
 CC -!- SUBCELLULAR LOCATION: Nuclear. Associated with mitotic
 CC chromosomes.
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN PYRAMIDAL CELLS IN JUVENILE
 CC HIPPOCAMPUS, GRANULAR CELLS IN JUVENILE CEREBELLAR CORTEX AND IN
 CC ADULT SPLEEN.
 CC -!- SIMILARITY: Belongs to the kinesin-like protein family.
 CC Chromokinesin subfamily.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch)

 DR EMBL; D12646; BAA02167.1; -.
 DR PIR; A54803; A54803.
 DR HSSP; P17119; 3KAR.
 DR MGD; MGI:108389; Kif4.
 DR InterPro; IPR001752; kinesin_motor.
 DR Pfam; PF00225; kinesin; 1.
 DR PRINTS; PRO0380; KINESINHEAVY.
 DR SMART; SM00129; KISC; 1.
 DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
 DR PROSITE; PS50067; KINESIN MOTOR DOMAIN2; 1.
 KW Motor protein; Microtubule; ATP-binding; DNA-binding;

KW Nuclear protein; Coiled coil.
 FT DOMAIN 1 350 KINESIN-MOTOR.
 FT DOMAIN 351 1000 COILED COIL (BY SIMILARITY).
 FT DOMAIN 1001 1231 GLOBULAR.
 FT NP BIND 88 95 ATP (POTENTIAL).
 FT CONFLICT 112 112 I -> S (IN REF. 2).
 SQ SEQUENCE 1231 AA; 139551 MW; F34F2C2D21158FE4 CRC64;

Query Match 28.6%; Score 710.5; DB 1; Length 1231;
 Best Local Similarity 35.8%; Pred. No. 9.5e-34;
 Matches 186; Conservative 81; Mismatches 184; Indels 69; Gaps 14;

QY 6 VAVCVRVRLNRSRESLG-ETAVYWKTDNNVIYQVDSKSFNDFRVFHGNETTKNYEE 64
 Db 10 VRVALRCRPLVSKKEIKGCQOTCLSFVPGPVV--VGNDKSFYDFVDFPSTREQEEVNT 67
 QY 65 IAAPIIDSAIQYNGTIFAYGQTASGKTYTMGS---EDH---LGVIPRAIHDIQKIK 117
 Db 68 AVAPLIKGVFGYNATVLAYGQTSGKTYSMGAYTAEGHSAIGVIFRVIQLLFRIN 127
 QY 118 KFPDFRFLRVSYMEIYNTITDLCGT-OKMKPLIREDVNRNVYVADLTVEVYVTSEM 176
 Db 128 KKSDFEFTLKVSYLEIYNEEILLCLCSSREKATQINREDPKGKIVGLTETKTVLVASD 187
 QY 177 ALKWITKGEKSHYGETKMNQSSSHTIFRMLESREKGEPSNCEGSKVSHNLVDLA 236
 Db 188 TVSCLEQGNNSTVASTAMNQSRSRSHAFTISIEQRKK---NDKNSFRSKLHLVDLA 243
 QY 237 GSERAAQTGAAGVRLKEGNCINRSLFILQVKKLSGQGVGFVNYRDSKLTILQNSLG 296
 Db 244 GSERQKTKAEGDRLREGINRGLLGLNGLSALGDKKGNFVYRDSKLTILQDSLG 303
 QY 297 GNPKTRITCTTPV--SPDETTLTALQFSTAKYMKNTPVYNEVSTDEALLKRYRKEIMDL 354
 Db 304 GNSHTLMIACVSPADSNLEETLTLYRADRKIKNKPIIN-----IDPQAAELNHL 355
 QY 355 KKOLEEVSL-----ETRAQAMEKQOLAQLLEBKLLQKVQNEKIN 395
 Db 356 KQVQOQLILLQAHGTLPGDINVEPSENLSLMEKNQ--SLVEENKLSRGLSEAAQ 413
 QY 396 ITRML---VTSSSLTLOELKAKRRRVTCWLGKINK---MKNSNYADQFNIPNTITK 448
 Db 414 TAQMLERILTEQANEKKNKLEELRRHACKVDQLKLVLEDDQELKENIELICNLQOV 473
 QY 449 THKLS-----INLREIDESVCS-----ESDVFS 472
 Db 474 IATLSDEAAACMTATIDTAGADTQVQSSPDTSRSSDVFS 513

RESULT 11
 KI22_STRPU
 ID KI22_STRPU STANDARD; PRT; 699 AA.
 AC P46872;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Kinesin-11 85 kDa subunit (KRP-85/95 85 kDa subunit).
 GN KRP85.
 OS Strongylocentrotus purpuratus (Purple sea urchin).
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
 OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
 OC Strongylocentrotus.
 OX NCBI_TaxID=7668;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Egg;
 RX MEDLINE=94050179; PubMed=8232586;
 RA Cole D.G., Chinn S.W., Wedaman K.P., Hall K., Vuong T.,
 RA Scholey J.M.;
 RT "Novel heterotrimeric kinesin-related protein purified from sea
 RT urchin eggs.";
 RL Nature 366:268-270(1993).
 CC -!- SUBUNIT: Heterotrimer of a 115 kDa subunit (KAP115) and two

CC kinesin-like subunits of 95 kDa (KRP95) and 85 kDa (KRP85).
CC -!- PTM: The N-terminus is blocked.
CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
CC II SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L16993; AAA16098.1; --
CC PIR: S38982; S38982.
CC HSP: P17119; 3KAR.
CC InterPro: IPR001752; kinesin_motor.
CC Pfam: PF00225; kinesin; 1.
CC PRINTS: PR00380; KINESINHEAVY.
CC SMART: SM00129; KISC; 1.
CC PROSITE: PS00411; KINESIN MOTOR DOMAIN1; 1.
CC PROSITE: PS0067; KINESIN MOTOR DOMAIN2; 1.
CC Motor protein; Microtubule; ATP-binding; Coiled coil.
CC FT DOMAIN 1 340 KINESIN-MOTOR (BY SIMILARITY).
CC COILED COIL (BY SIMILARITY).
CC FT DOMAIN 341 619
CC GLOBULAR (BY SIMILARITY).
CC FT NP_BIND 620 699
CC ATP (POTENTIAL).
CC FT NP_BIND 97 104
CC SEQUENCE 699 AA; 78697 MW; 783866111CB08190 CRC64;
CC SQ

Query Match 28.4%; Score 704.5; DB 1; Length 699;
Best Local Similarity 36.2%; Pred. No. 1e-33;
Matches 200; Conservative 76; Mismatches 152; Indels 125; Gaps 19;
QY 6 VAVCVVRPLNSRESIG-----ETAAQVYWKTDNNVIVQDGSKFNEDRVPHG 54
DB 11 VRVVRVCRPLNSKETGQGFVSKVWKMDEMGRGVQV---TNPNA-PSGEPKSFETDTPFAP 66
QY 55 NETTKNYVEETAAPIDSAIGYNGTIFAYGOTASGTYTWMG--SEDLH-GVIPRAIHD 111
DB 67 GAKQTDVNVQTPARPLVDAILIIGYNGTIFAYGOTGTGKFTWEGVRSQPELAGIPNSFAH 126
QY 112 IFQKIKKFPDR-EPFLRVSYMEIYNETITDLCGQPKMKPLIREDVNRNVYVADLITEV 170
DB 127 IFGHIAKQENVRFLRVSVYLEIYNEEVKDLL-GRDQQRLEVRKRPDVGVVVVDLSAFV 185
QY 171 VYTEMALKWITKGSRHYGETKQNRSSRSHITFRMILSRKSGPSCNCEGSKVQVSHL 230
DB 186 VNNADMDRIMTGLNKNRSGVATNNSSSRSHAFITTLERSDMG--LDKEQHVVRGKL 243
QY 231 NIVDLAGSERAQTCGAAGVRLKEGNCINRSIFILQVTKKLSDGQVGGFINYRDSKLTRI 290
DB 244 HMVDLAGSERQTKTGATGRLKEATKINLSLTLGNVLSLVLDGK-STHIPPVRSKLTRL 302
QY 291 LQNSLGGNPKTRIICITIPV--SPDETITAFQFASTAKYMKNPVNVESVDEALLKRYR 348
DB 303 LQDSLGGNAKTVWCANIGPAENYDETITSLRYANRAKNIKNKAKINEDPKD-ALLREBFQ 361
QY 349 KEIMDLKKQLEF-----VSLETRAQAM 370
DB 362 KEIEELKXQISSEGLDDDEBSGSEGEAGEGGVKKKKNKKNPKKLSPELMA-AM 420
QY 371 EK--DQLAQLEE-KOLL-----QKVNQKNIENLTMLVT- 402
DB 421 QKKIDEEKALKREKMDVREDNTVHRELQRSELSLHKAQDDQKIENKLNIAIKKLIIVG 480
QY 403 -----SSSLITLQQLKAKRKRVTWGLKINMKNSNVADQFNPTNIT 446
DB 481 GVDLLAKSBEQQLLEQSALENKERMKAQESMR-----KNMEEREQERMDEEKYS 531
QY 447 T-----KTHKL 452
DB 532 SLQDEAHGKTCKL 544

RESULT 12
KL61 DROME
ID KL61 DROME STANDARD; PRT; 1066 AA.
AC P46863; Q8TOA6; Q9W018;
DT 01-NOV-1995 (Rel. 32, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Bipolar kinesin KRP-130 (Kinesin-like protein Klp61F).
GN KLP61F OR KLP2 OR CG9191.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=94043448; PubMed=8227131;
RA Heck M.M.S., Pereira A., Pesavento P.A., Yannoni Y., Spradling A.C.,
RA Goldstein L.S.B.;
RT "The kinesin-like protein KLP61F is essential for mitosis in
RT Drosophila.";
RL J. Cell Biol. 123:665-679 (1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harsis N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Strong R., Sun E.,
RA Spier E., Spradling A.C., Stapleton M., Venter E., Wang A.H., Wang X.,
RA Svirskaas R., Tector C., Turner R., Venter E., Weissbach J.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195 (2000).
RN [3]
RP REVISIONS.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.B.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,

QY 411 ELKAKRRVTVWCLGKINKMKNVADQ---FNIPNTIT-----TKTHKLSINLLREI 460
 Db 401 DTEAEKQLIREYERLARKADYAEQSRVQLQEDITAMRNSYDKLSTLQENLRKEK 460
 QY 461 DESVCSSESVFNTLTLSEIE 482
 Db 461 ETEAILKAEVLCKT-EVMSRAE 481

RESULT 14
 KF4A_XENLA STANDARD; PRT; 1226 AA.
 ID KF4A_XENLA Q9PSI0;
 AC Q91784; Q9PSI0;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Chromosome-associated kinesin KLP1 (Chromokinesin).
 GN KLP1.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
 OC Xenopodinae; Xenopus.
 OC NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, AND TISSUE SPECIFICITY.
 RC TISSUE=Oocyte;
 RX MEDLINE=95236444; PubMed=7720067;
 RA Vernos I., Raats J., Hirano T., Heasman J., Karsenti E., Wylie C.;
 RT "Xklp1, a chromosomal Xenopus kinesin-like protein essential for spindle organization and chromosome positioning.";
 RL Cell 81:117-127(1995).
 RN [2]
 RP SEQUENCE OF 9-338 FROM N.A.
 RX MEDLINE=93246065; PubMed=8482413;
 RA Vernos I., Heasman J., Wylie C.;
 RT "Multiple kinesin-like transcripts in Xenopus oocytes.";
 RL Dev. Biol. 157:232-239(1993)
 CC -!- FUNCTION: REQUIRED FOR MITOTIC CHROMOSOMAL POSITIONING AND BIPOLAR SPINDLE STABILIZATION.
 CC -!- SUBCELLULAR LOCATION: Nuclear. Associated with mitotic chromosomes.
 CC -!- TISSUE SPECIFICITY: Expressed in oocytes, eggs, testes and brain.
 CC -!- SIMILARITY: Belongs to the kinesin-like protein family. Chromokinesin subfamily.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

 CC EMBL; X82012; CAA57539.1; -.
 DR PIR; I51617; I51617.
 DR HSSP; P17119; 3KAR.
 DR InterPro; IPR001752; kinesin_motor.
 DR Pfam; PF00225; kinesin; 1.
 DR PRINTS; PR00380; KINESINHEAVY.
 DR SMART; SM00129; KISC; 1.
 DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
 DR PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
 DR Motor protein; Microtubule; ATP-binding; DNA-binding;
 KW Nuclear protein; Coiled coil.
 FT DOMAIN 1 350 KINESIN-MOTOR.
 FT DOMAIN 351 1006 COILED COIL (BY SIMILARITY).
 FT DOMAIN 1007 1226 GLOBULAR.
 FT NP_BIND 87 94 ATP (POTENTIAL).
 FT CONFLICT 163 163 I -> L (IN REF. 2).
 FT SEQUENCE 1226 AA; 138923 MW; 7F0275FCF3316697 CRC64;

RX MEDLINE=20370907; PubMed=10846156;
 RA Setou M., Nakagawa T., Seog D.-H., Hirokawa N.;
 RT "Kinesin superfamily motor protein KIF17 and mLin-10 in NMDA receptor-containing vesicle transport.";
 RL Science 288:1796-1802(2000).
 CC -!- FUNCTION: Transports vesicles containing N-methyl-D-aspartate (NMDA) receptor 2B along microtubules.
 CC -!- SUBUNIT: Interacts with LIN-10 PDZ domain.
 CC -!- TISSUE SPECIFICITY: Neuronal-specific.
 CC -!- SIMILARITY: Belongs to the kinesin-like protein family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL; AB008867; BAB21099.1; -.
 DR HSSP; P17119; 3KAR.
 DR MGD; MGI:1098229; Kif17.
 DR GO; GO:0005871; C:kinesin complex; IDA.
 DR GO; GO:0003777; F:microtubule motor activity; IDA.
 DR GO; GO:0005515; F:protein binding; IPI.
 DR GO; GO:0005515; F:protein binding; IPI.
 DR GO; GO:0007017; P:microtubule-based process; IDA.
 DR GO; GO:0016192; P:vesicle-mediated transport; IDA.
 DR InterPro; IPR001752; kinesin_motor.
 DR Pfam; PF00225; kinesin; 1.
 DR PRINTS; PR00380; KINESINHEAVY.
 DR SMART; SM00129; KISC; 1.
 DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
 DR PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
 DR Motor protein; Microtubule; ATP-binding; Coiled coil; Transport;
 KW Protein transport.
 FT DOMAIN 1 265 KINESIN-MOTOR.
 FT DOMAIN 346 470 COILED COIL (POTENTIAL).
 FT DOMAIN 748 855 COILED COIL (POTENTIAL).
 FT NP_BIND 91 98 ATP (POTENTIAL).
 FT SEQUENCE 1038 AA; 116372 MW; 2BED852A3AFDBD46 CRC64;

Query Match 28.3%; Score 702.5; DB 1; Length 1038;
 Best local similarity 37.1%; Pred. No. 2,2e-33;
 Matches 186; Conservative 88; Mismatches 179; Indels 49; Gaps 15;

QY 5 AVAVCVRVPLNLSREESLG-ETFAQVYVWKTDDNNVIVQDGS-----KSFNFRVFGHNETT 58
 Db 5 SVKVVRCRPMKRRERELSCQSVTVVDSARGQCFTQNGAABDPKQFTFDGAYIEHFT 64
 QY 59 KNYVEIARPIIDSATQYNGTIFAYGQTASGKTYTMGSED---HLGVIPRAIHDFQK 115
 Db 65 EQIYNEIAYPLVEGVTEGNTIFAYGQTGSKSFTMQGLPDPCCQGIIPRAFEHVFS 124
 QY 116 IKKFPDFELLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNVYVADLTVEVVTSE 175
 Db 125 VQCAENTKFLVASYLEIYNEDVHDL-LGADTQRLKELKHPKGVYVKGSLMHTVHYVA 183
 QY 176 MALKWTITGEKSHYGETKQNRSSRSHITFRMLE---SREKGEFSNCEGSYKVSHLNL 232
 Db 184 QCBVMTGWKRAVGYTLNKKDSSRSHISFTINIEIYAVDERG-----KDLHRAKGLNL 238
 QY 233 VDLAGSRAAQTCAAGVRKKEGNCINRSLFILGVTKLSDGQGVGFYNRDLSKLTRLQ 292
 Db 239 VDLAGSRQSKTGATGRLKEATKINLSALGNVISALVDGRC-KHIPYRDSKLTRLQ 297
 QY 293 NSLGGNPKTRIICTIPV--SPDETALQFASTAKYMKVTPPVNEVSTDEALKRYRKE 350
 Db 298 DSLGGNTKTLWVACLSPADNNYDETTLTRYANRAKNKPRINPKD-ALLREYQEE 356
 QY 351 IMDLKKOLEVSLQETRAQAMEKQDLAQLEEKDLLOKVQNFKLENLRLMTVSSSLTLOQ 410
 Db 357 IKRLKAIL-----AQQMGPGNLSALLSTQTTPPGVPVQSEE-----KLL---SPTTVQ 400

Query Match	27.9%;	Score 692;	DB 1;	Length 1226;
Best Local Similarity	38.7%;	Pred. No. 1.1e-32;		
Matches 182;	Conservative 70;	Mismatches 148;	Indels 70;	Gaps 15;
QY	2	BEG-AVAVCVRVRLPNSREESLG-ETAQVYKTNVNIYQVDGSGKSNFDRVFGHNETTK	59	
Db	4	DEGIPVRVALRCRPLVPKENNEGKCMCLTFVPEGQQVI--VGTEKSFYDYVDFPSAEQE	61	
QY	60	NVYBEIAAPIIDSALQGVNGTIFAYGQTASCTKTYMGSDEH-----LQVIFRAHDI	112	
Db	62	EVVNSAVAPLIKGLFKGNVNLVAYQGTSGKTYSMGGAYTHNOENFTVGVPIRTVIAL	121	
QY	113	FQKTKPPDREFLLRVSYMEIYNTITDLL-CGTQMKPLIIRDVNRNVVAALTEVV	171	
Db	122	FREIHQPEWENFLKVSYLEIYNEEILDLLVAARDKNTISIRBDPEKGKICGLTERDV	181	
QY	172	YTSEMALKWTIKGKSRHYGHTQONQRSSHTTIFRMILESRKEGSPNCEGSKVKSHLN	231	
Db	182	KTALDITLSCLEQGNSSRTVASTAMNSQSSRSHAFTTISIEQRKEGDKNN--SFR-SKLH	237	
QY	232	LVDLAGSERAATGAAGVRUKGECNVINRSILFILQVITKILSD-GVQGGFINYRDRSKLTRI	290	
Db	238	LVDLAGSEROKTKVAGDRUKGEGISINRGLLCGLGNVISALGDESKGGGFYFDRSKLTRL	297	
QY	291	LQNSLGNPNKTRILICTITPV--SPDETILAQFASTAKYMKNTPYNVNEVSTDDALLKRYR	348	
Db	298	LQDSLGCNSHTLMIACVSPADSNMEETLNTIRYADRARKINKEPIVN-TDQAAELQORLK	356	
QY	349	KEIMDLKKQLQEEV-----SLETRAQAMEKD-----QLAQ	377	
Db	357	LQVQELQVLLQAHGGTLPVLNSMEPSENLQSLMERNKNLEKENGKLSRELGEAAVQTQA	416	
QY	378	LLBEKDLILQVQNEKI-----ENLTRMLVTSSSITLQOELK	413	
Db	417	FL-EKIMTQEQNEKLGSKNMELKQHAACKVNTQLRVETLE----DQELK	461	

```

RESULT 15
KF4A_CHICK
ID KF4A_CHICK STANDARD; PRT; 1225 AA.
AC Q90640; Q90608;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chromosome-associated kinesin KIF4A (Chromokinesin).
GN KIF4A.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, AND TISSUE
RP SPECIFICITY.
RC STRAIN=White leghorn; TISSUE=Embryonic retina;
RC MEDLINE=95181533; PubMed=7876303;
RA Wang S.Z., Adler R.;
RT "Chromokinesin: a DNA-binding, kinesin-like nuclear protein.";
RT J. Cell Biol. 128:761-768(1995).
RN [2]
RP SEQUENCE OF 728-1088 FROM N.A.
RP STRAIN=White leghorn; TISSUE=Embryonic retina;
RC MEDLINE=94151328; PubMed=8108415;
RA Wang S.Z., Adler R.;
RT "A developmentally regulated basic-leucine zipper-like gene and its
RT expression in embryonic retina and lens.";
RT Proc. Natl. Acad. Sci. U.S.A. 91:1351-1355(1994).
CC -!- FUNCTION: REQUIRED FOR MITOTIC CHROMOSOMAL POSITIONING AND BIPOLAR
CC SPINDLE STABILIZATION.
CC -!- SUBCELLULAR LOCATION: Nuclear. Associated with mitotic
CC chromosomes.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN PROLIFERATING CELLS;
CC NEUROEPITHELIUM OF EMBRYOS.
CC

```

Search completed: July 29, 2004, 09:36:17
Job time : 5.67769 secs

us-10-045-631b-88_copy_2_488.rsp

Fri Aug 6 10:49:14 2004

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 09:27:31 ; Search time 20.6858 Seconds

(without alignments)
7428.155 Million cell updates/sec

Title: US-10-045-631b-88_COPY_2_488

Perfect score: 2483

Sequence: 1 AEGAVAVCVVRPLNSREE.....SDVFSNTLDTSEIENWNPAT 487

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mmc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteria.*
- 17: sp_archaea.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2021.5	81.4	549	11 Q7TPX4	Q7tpx4 mus musculus
2	1584	63.8	2954	13 Q42263	Q42263 xenopus lae
3	880	35.4	807	10 Q94HV9	Q94hv9 arabidopsis
4	880	35.4	823	10 Q9S7P3	Q9s7p3 arabidopsis
5	856	34.5	459	10 Q9S830	Q9s830 arabidopsis
6	844.5	34.0	888	10 Q9L062	Q9l062 arabidopsis
7	801	32.3	1885	5 Q869B8	Q869b8 dictyosteli
8	780.5	31.4	2013	5 Q9VK10	Q9vk10 drosophila
9	780.5	31.4	2244	5 Q9NCG0	Q9ncg0 drosophila
10	769.5	31.0	1055	10 Q8RWM4	Q8rwm4 arabidopsis
11	769.5	31.0	1055	10 Q8WSR5	Q8wsr5 arabidopsis
12	759	30.6	1459	3 P87198	P87198 ustilago ma
13	753	30.3	1058	10 Q9SJU0	Q9sjuo arabidopsis
14	748.5	30.1	1033	10 Q9LHL9	Q9lhl9 arabidopsis
15	737.5	29.7	699	5 Q8MPT8	Q8mpt8 caenorhabdi
16	735.5	29.6	747	11 Q8BNH4	Q8bnh4 mus musculus

17	735.5	29.6	757	11 Q8OU27	Q8ou27 mus musculu
18	733.5	29.5	443	11 Q8CGJ1	Q8cgj1 mus musculu
19	729.5	29.4	890	10 Q8WSR6	Q8wsr6 arabidopsis
20	728	29.3	701	11 Q7TSZ7	Q7tsz7 mus musculu
21	727.5	29.3	956	10 Q9C7B9	Q9c7b9 arabidopsis
22	727	29.3	997	10 Q9FG03	Q9fg03 arabidopsis
23	725	29.2	782	5 Q19633	Q19633 caenorhabdi
24	725	29.2	782	5 Q9GV92	Q9gv92 caenorhabdi
25	724	29.2	959	10 Q8S950	Q8s950 nicotiana t
26	720.5	29.0	735	5 Q9U0D5	Q9u0d5 tetrahymena
27	720	29.0	408	11 Q8OUK1	Q8ouk1 mus musculu
28	717.5	28.9	954	10 Q9AMW8	Q9awm8 oryza sativ
29	717.5	28.9	1193	10 Q7X7H8	Q7x7h8 oryza sativ
30	716.5	28.9	408	4 Q81WH8	Q81wh8 homo sapien
31	716.5	28.9	677	5 Q9VRK9	Q9vrk9 drosophila
32	716.5	28.9	699	13 Q98T11	Q98t11 xenopus lae
33	714	28.8	836	5 Q7YUC7	Q7yuc7 tetrahymena
34	713.5	28.7	1127	4 Q86TN3	Q86tn3 homo sapien
35	712.5	28.7	671	4 Q86XX7	Q86xx7 homo sapien
36	710.5	28.6	819	11 Q7TQG6	Q7tqg6 mus musculu
37	710.5	28.6	974	10 Q8S905	Q8s905 arabidopsis
38	710.5	28.6	1231	11 Q80YP3	Q80yp3 mus musculu
39	707	28.5	160	11 Q35059	Q35059 mus musculu
40	706.5	28.5	744	13 Q93478	Q93478 xenopus lae
41	705.5	28.4	1037	10 Q3FH58	Q3fh58 arabidopsis
42	705	28.4	1121	10 Q8SVI8	Q8svi8 arabidopsis
43	702.5	28.3	671	5 Q8MPT7	Q8mpt7 caenorhabdi
44	702.5	28.3	672	5 Q9GV93	Q9gv93 caenorhabdi
45	700	28.2	909	10 Q9XI03	Q9xi03 arabidopsis

ALIGNMENTS

RESULT 1

Q7TPX4 PRELIMINARY; PRT; 549 AA.

AC Q7TPX4;
DT 01-OCT-2003 (TREMBLrel. 25, Created)
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Egg;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.D., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S., Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E., Jones S.J., Marra M.A.;
"Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences";
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Egg;

RA Strausberg R.;
 RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC052843; AAH52843.1; --
 KW Hypothetical protein.
 FT NON TER 549 549
 SQ SEQUENCE 549 AA; 62899 MW; B707C97DC9FB38D3 CRC64;

Query Match 81.4%; Score 2021.5; DB 11; Length 549;
 Best Local Similarity 80.3%; Pred. No. 1.7e-116;
 Matches 396; Conservative 35; Mismatches 53; Indels 9; Gaps 2;

QY 1 AEEGAVAVCVVRPLNSRESLGETAQVYKMTDNNVIYQVDGSKSFNDFRVFHNETHKN 60
 DB 1 AEEGAVAVCVVRPLNSRESLGETAQVYKMTDNNVIYQVDGSKSFNDFRVFHNETHKN 60
 QY 61 VYEEIAAPIIDSAIQNGYNGTIFAYGQTASGKTYTMMGSEDLGVIPRAIHDIQKIKFP 120
 DB 61 VYEEIAAPIIDSAIQNGYNGTIFAYGQTASGKTYTMMGSEDLGVIPRAIHDIQKIKFP 120
 QY 62 VYEEIAVPIISSAIIQNGYNGTIFAYGQTASGKTYTMMGSEDLGVIPRAIHDIQKIKFP 121
 DB 62 VYEEIAVPIISSAIIQNGYNGTIFAYGQTASGKTYTMMGSEDLGVIPRAIHDIQKIKFP 121
 QY 121 DREPLLRVSMYIYNETITDLCGTOKMKPLIREDVNRVYVADLTVEEVYTTSEMAKW 180
 DB 121 DREPLLRVSMYIYNETITDLCGTOKMKPLIREDVNRVYVADLTVEEVYTTSEMAKW 180
 QY 122 ERBFLLRVSMYIYNETITDLCNAQKKKPLIREDNRYVYVADLTVEEVYTTSEMAKW 181
 DB 122 ERBFLLRVSMYIYNETITDLCNAQKKKPLIREDNRYVYVADLTVEEVYTTSEMAKW 181
 QY 181 ITKGEKSRHYGETKMNQSRSSHTIFRMILESREKGEPSNCEGSKVSHLNLVDLAGSER 240
 DB 181 ITKGEKSRHYGETKMNQSRSSHTIFRMILESREKGEPSNCEGSKVSHLNLVDLAGSER 241
 QY 241 AAGTGAAGVRLKEGCNINRSILFGLGVIKLSDGVGFINYRDSKLTILQNSLGNPK 300
 DB 241 AAGTGAAGVRLKEGCNINRSILFGLGVIKLSDGVGFINYRDSKLTILQNSLGNPK 301
 QY 301 TRICTITPVSFDETLTALQAEKLLQVQNEKIENLRLMVLTSSTLTQLQELKQLEE 360
 DB 301 TRICTITPVSFDETLTALQAEKLLQVQNEKIENLRLMVLTSSTLTQLQELKQLEE 361
 QY 361 VSLTEQAQMEKQQLAQLEKLLQVQNEKIENLRLMVLTSSTLTQLQELKQKRRRV 420
 DB 361 VSLTEQAQMEKQQLAQLEKLLQVQNEKIENLRLMVLTSSTLTQLQELKQKRRRV 421
 QY 421 TWLGKINKMKNVADQFNPTITTKLKSINLRLI-----DESVCSESDVFSNT 474
 DB 421 TWLGKINKMKNVADQFNPTITTKLKSINLRLI-----DESVCSESDVFSNT 475
 QY 475 LDTLSELEWNPAT 487
 DB 475 LDTLSELEWNPAT 487
 QY 479 LESLAEVWSSAT 491
 DB 479 LESLAEVWSSAT 491

RESULT 2
 ID C42263 PRELIMINARY; PRT; 2954 AA.
 AC C42263;
 DT 01-JAN-1998 (TREMBlrel. 05, Created)
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Kinesin-related protein.
 GN XCENP-E. laevis (African clawed frog).
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
 OC Xenopodinae; Xenopus.
 OC NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98028574; PubMed=9363944;
 RA Wood K.W., Sakowicz R., Goldstein L.S., Cleveland D.W.;
 RT "CENP-E is a plus end-directed kinetochore motor required for
 metaphase chromosome alignment."
 RL Cell 91:357-366(1997)
 DR EMBL; AF027728; AAC60300.1; --
 DR F01; T14156; T14156.
 DR HSSP; P33176; 1BG2.
 DR GO; GO:0005871; C:kinesin complex; IEA.

DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0003774; F:Motor activity; IEA.
 DR GO; GO:0007017; F:Microtubule-based process; IEA.
 DR InterPro; IPR001752; kinesin_motor.
 DR Pfam; PF00225; kinesin; 1.
 DR PRINTS; PR00380; KINESINHEAVY.
 DR SMART; SM00129; KISC; 1.
 DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
 DR PROSITE; PS00657; KINESIN MOTOR DOMAIN2; 1.
 DR PROSITE; PS00657; KINESIN MOTOR DOMAIN2; 1.
 SQ SEQUENCE 2954 AA; 339964 MW; 439804ED0E592679 CRC64;

Query Match 63.8%; Score 1584; DB 13; Length 2954;
 Best Local Similarity 64.0%; Pred. No. 1.4e-88;
 Matches 320; Conservative 68; Mismatches 96; Indels 16; Gaps 6;

QY 1 AEEGAVAVCVVRPLNSRESLGETAQVYKMTDNNVIYQVDGSKSFNDFRVFHNETHKN 60
 DB 1 AEEGAVAVCVVRPLNSRESLGETAQVYKMTDNNVIYQVDGSKSFNDFRVFHNETHKN 60
 QY 61 VYEEIAAPIIDSAIQNGYNGTIFAYGQTASGKTYTMMGSEDLGVIPRAIHDIQKIKFP 120
 DB 61 VYEEIAAPIIDSAIQNGYNGTIFAYGQTASGKTYTMMGSEDLGVIPRAIHDIQKIKFP 120
 QY 60 IYGEIAVPIIRSAIQNGYNGTIFAYGQTASGKTYTMMGSEDLGVIPRAIHDIQKIKFP 119
 DB 60 IYGEIAVPIIRSAIQNGYNGTIFAYGQTASGKTYTMMGSEDLGVIPRAIHDIQKIKFP 119
 QY 121 DREPLLRVSMYIYNETITDLCGTOKMKPLIREDVNRVYVADLTVEEVYTTSEMAKW 180
 DB 121 DREPLLRVSMYIYNETITDLCGTOKMKPLIREDVNRVYVADLTVEEVYTTSEMAKW 180
 QY 120 NRBFLLRVSMYIYNETITDLCDDRKKKPLIREDNRYVYVADLTVEEVYTTSEMAKW 179
 DB 120 NRBFLLRVSMYIYNETITDLCDDRKKKPLIREDNRYVYVADLTVEEVYTTSEMAKW 179
 QY 181 ITKGEKSRHYGETKMNQSRSSHTIFRMILESREKGEPSNCEGSKVSHLNLVDLAG 237
 DB 181 ITKGEKSRHYGETKMNQSRSSHTIFRMILESREKGEPSNCEGSKVSHLNLVDLAG 239
 QY 238 SERAAGTGAAGVRLKEGCNINRSILFGLGVIKLSDGVGFINYRDSKLTILQNSLGG 297
 DB 238 SERAAGTGAAGVRLKEGCNINRSILFGLGVIKLSDGVGFINYRDSKLTILQNSLGG 299
 QY 240 SERASQTGAAGVRLKEGCNINRSILFGLGVIKLSDGVGFINYRDSKLTILQNSLGG 357
 DB 240 SERASQTGAAGVRLKEGCNINRSILFGLGVIKLSDGVGFINYRDSKLTILQNSLGG 359
 QY 298 NPKTRICTITPVSFDETLTALQAEKLLQVQNEKIENLRLMVLTSSTLTQLQELKQ 415
 DB 298 NPKTRICTITPVSFDETLTALQAEKLLQVQNEKIENLRLMVLTSSTLTQLQELKQ 418
 QY 300 NAKTVIITPVSFDETLTALQAEKLLQVQNEKIENLRLMVLTSSTLTQLQELKQ 471
 DB 300 NAKTVIITPVSFDETLTALQAEKLLQVQNEKIENLRLMVLTSSTLTQLQELKQ 471
 QY 358 LE--EVSLETRAQAMEKQQLAQLEKLLQVQNEKIENLRLMVLTSSTLTQLQELKQ 471
 DB 358 LE--EVSLETRAQAMEKQQLAQLEKLLQVQNEKIENLRLMVLTSSTLTQLQELKQ 471
 QY 416 RKRVTWCLGKINKMKNVADQFNPTITTKLKSINLRLI-----DESVCSESDVFSNT 471
 DB 416 RKRVTWCLGKINKMKNVADQFNPTITTKLKSINLRLI-----DESVCSESDVFSNT 471
 QY 419 RKRVTWAPGKIQNSLHSGVSDFDMLSRGPNFSKXKAKFSDMPSFPEIDDSVCTEFSDF 478
 DB 419 RKRVTWAPGKIQNSLHSGVSDFDMLSRGPNFSKXKAKFSDMPSFPEIDDSVCTEFSDF 478
 QY 472 SNTLDTLS---EIEWNPAT 487
 DB 472 SNTLDTLS---EIEWNPAT 487
 QY 479 DDALSMDSNGIDAEWNLAS 498
 DB 479 DDALSMDSNGIDAEWNLAS 498

RESULT 3
 ID Q94HV9 PRELIMINARY; PRT; 807 AA.
 AC Q94HV9;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Kinesin motor protein (kin2), putative.
 GN T4M14.11.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Town C.D., Haas B.J., Wu D., Maiti R., Hannick L.I., Chan A.P.,
 RA Tallon L.J., Rooney T., Utterback T.R., VanAken S.E., Feldblum T.V.,
 RA White O., Fraser C.M.;
 RA "Arabidopsis thaliana chromosome 1 BAC T4M14 genomic sequence.";

```

RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC027036; AAK6792.1; -.
DR GO; GO:0005871; C:kinesin complex; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR GO; GO:0007017; F:microtubule-based process; IEA.
DR InterPro; IPR001752; kinesin_motor.
DR Pfam; PF00225; kinesin; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
DR PROSITE; PS0067; KINESIN MOTOR DOMAIN2; 1.
SQ SEQUENCE 807 AA; 91260 MW; FD7CDAD68EA30C28 CRC64;

Query Match 35.4%; Score 880; DB 10; Length 807;
Best Local Similarity 43.1%; Pred. No. 6.1e-46;
Matches 220; Conservative 70; Mismatches 166; Indels 54; Gaps 12;

QY 6 VAVCVVRPLNSRRESLSGETAQVYKTDNNVIYQVDSKSFN-----FDRVPHGNE 56
DQ : ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4 ICVAVRVRP-----PAPENGASLWKVEDN---RISLHKSLODPTTASHADHVFDESS 54

QY 57 TTKNYVEEIAPIIDSAIQGYNGTIFAYQOTASGKTYTMGSEDLHGVIPRAIHDFQKI 116
DQ : ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 55 TNASVYELLTKDIIHAAVEGFNGTAFAYQOTSGKTFMTGSETDPIIRRSVRDVFERI 114

QY 117 KFPDREFLLRVSMEIYNEITDILCGTQKMKPLIREDVNRNVTVDLTSEVVTSEM 176
DQ : ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 115 HMSIDREFLIRVSMEIYNEEINDLL--AVENQRQIHEHLRGVFAGLKEEIVSDAQ 172

QY 177 ALKWTITKEGRHYGETKMNQSRSHSTIFRMILESRKEGPECSGSKVSHLNLVDLA 236
DQ : ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 173 ILKLDSGEVNRHFGETNMNVHSSRSHSTIFRMVIESR--GKDNSSSDAIRSVMLNVDLA 230

QY 237 GSERAAQTGAAGVRLKEGCNINRSFILGQVTKKLSG--QVGGFINYRDSKLTILQNSL 295
DQ : ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 231 GSERIAKTGAGVRLQEGKYINKSLMILGNVINKLSDTKLAHPIPYRDSKLTILQPAL 290

QY 296 GGNPKTRICTITPVS--FDETLTALQPASTAKYMKNTPYNEVSTDEALLKRYKEIMD 353
DQ : ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 291 GGNAKTCTICTTAPEHHIEESKGTLOFASRAKRTNCAQVNEILTDAALLKROKLEIEE 350

QY 354 LKKOLEVSLTRAQ-----AMEKDQLAQLLEEKDLQKQVNE-----KIE 394
DQ : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 351 LRMLKGSHAEVLEQEILNLSNOMLYELECERLKTQLEEEKRQKEQENCICEQOMKIE 410

QY 395 NLTRMLVTSSSLTQOELKAKRRVTVCLGKINKMKNKNYADQNPITNITTKTKLSI 454
DQ : ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 411 NLNN-FVTNSDF-----KRNQSEDFIISRKTPDGLCNVNDTSDVPGTFCFKSASRSF 461

QY 455 NLLREIDESVCSSESVFNTL--DTLSEIEW 483
DQ : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 462 VVARSNVYSGLSDFSPMVHSLGDAVEDTW 491

RESULT 4
Q9S7P3 PRELIMINARY; PRT; 823 AA.
AC Q9S7P3;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Kinesin-like protein.
GN ZCFI25.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Ooe H., Kato A., Komeda Y.;

RT "Arabidopsis thaliana genomic sequence for a kinesin-like protein.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=20018182; PubMed=10548732;
RA Kato A., Suzuki M., Kuwahara A., Ooe H., Higano-Inaba K., Komeda Y.;
RT "Isolation and analysis of cDNA within a 300 kb Arabidopsis thaliana
RT genomic region located around the 100 map unit of chromosome 1.";
RL Gene 239:309-316(1999).
DR EMBL; AB028470; BAA88114.1; -.
DR EMBL; AB028468; BAA88112.1; -.
DR PIR; T52425; T52425.
DR HSSP; P33176; 1BG2.
DR GO; GO:0005871; C:kinesin complex; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR GO; GO:0007017; F:microtubule-based process; IEA.
DR InterPro; IPR001752; kinesin_motor.
DR Pfam; PF00225; kinesin; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
DR PROSITE; PS0067; KINESIN MOTOR DOMAIN2; 1.
SQ SEQUENCE 823 AA; 93148 MW; 6AFB1C622B4632C9 CRC64;

Query Match 35.4%; Score 880; DB 10; Length 823;
Best Local Similarity 43.1%; Pred. No. 6.3e-46;
Matches 220; Conservative 70; Mismatches 166; Indels 54; Gaps 12;

QY 6 VAVCVVRPLNSRRESLSGETAQVYKTDNNVIYQVDSKSFN-----FDRVPHGNE 56
DQ : ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4 ICVAVRVRP-----PAPENGASLWKVEDN---RISLHKSLODPTTASHADHVFDESS 54

QY 57 TTKNYVEEIAPIIDSAIQGYNGTIFAYQOTASGKTYTMGSEDLHGVIPRAIHDFQKI 116
DQ : ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 55 TNASVYELLTKDIIHAAVEGFNGTAFAYQOTSGKTFMTGSETDPIIRRSVRDVFERI 114

QY 117 KFPDREFLLRVSMEIYNEITDILCGTQKMKPLIREDVNRNVTVDLTSEVVTSEM 176
DQ : ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 115 HMSIDREFLIRVSMEIYNEEINDLL--AVENQRQIHEHLRGVFAGLKEEIVSDAQ 172

QY 177 ALKWTITKEGRHYGETKMNQSRSHSTIFRMILESRKEGPECSGSKVSHLNLVDLA 236
DQ : ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 173 ILKLDSGEVNRHFGETNMNVHSSRSHSTIFRMVIESR--GKDNSSSDAIRSVMLNVDLA 230

QY 237 GSERAAQTGAAGVRLKEGCNINRSFILGQVTKKLSG--QVGGFINYRDSKLTILQNSL 295
DQ : ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 231 GSERIAKTGAGVRLQEGKYINKSLMILGNVINKLSDTKLAHPIPYRDSKLTILQPAL 290

QY 296 GGNPKTRICTITPVS--FDETLTALQPASTAKYMKNTPYNEVSTDEALLKRYKEIMD 353
DQ : ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 291 GGNAKTCTICTTAPEHHIEESKGTLOFASRAKRTNCAQVNEILTDAALLKROKLEIEE 350

QY 354 LKKOLEVSLTRAQ-----AMEKDQLAQLLEEKDLQKQVNE-----KIE 394
DQ : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 351 LRMLKGSHAEVLEQEILNLSNOMLYELECERLKTQLEEEKRQKEQENCICEQOMKIE 410

QY 395 NLTRMLVTSSSLTQOELKAKRRVTVCLGKINKMKNKNYADQNPITNITTKTKLSI 454
DQ : ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 411 NLNN-FVTNSDF-----KRNQSEDFIISRKTPDGLCNVNDTSDVPGTFCFKSASRSF 461

QY 455 NLLREIDESVCSSESVFNTL--DTLSEIEW 483
DQ : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 462 VVARSNVYSGLSDFSPMVHSLGDAVEDTW 491

RESULT 5
Q9SS30 PRELIMINARY; PRT; 459 AA.
AC Q9SS30;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)

```

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative kinesin-like centromere protein.
GN F14P13.22
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,
RA Ronning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,
RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RA "Arabidopsis thaliana chromosome III BAC F14P13 genomic sequence.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC009400; AA02823.1; -;
DR HSP; P33176; 18G2.
DR GO; GO:0005871; C:kinesin complex; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR GO; GO:0007017; P:microtubule-based process; IEA.
DR InterPro; IPR001752; kinesin_motor.
DR Pfam; PF00225; Kinesin; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
DR PROSITE; PS00067; KINESIN MOTOR DOMAIN2; 1.
SQ SEQUENCE 459 AA; 51872 MW; 768487E9B7038E6A CRC64;
Query Match 34.5%; Score 856; DB 10; Length 459;
Best Local Similarity 50.7%; Pred. No. 8.8e-45;
Matches 194; Conservative 62; Mismatches 97; Indels 30; Gaps 8;
Qy 49 DRVPHGNETTKVVEETAAPIIDSAIQYNGTIFAYGCTAGSKTYTMGSDHLGVIPRA 108
Db 45 DRIFREDCKTVQVYEARTKEIVSAVRGNGTVFAYGQTNSGKTHMRGSPFEGVPLA 104
Qy 109 IHDIFQKIKFPDFREFLLRVSYMEIYNETITDLCCGTQKMKPLIREDVNRNVYADLTE 168
Db 105 VHDLFDTIYQDASREFLLRMSLYEIYNEIDNDLI--APEHRKLOIHENLEKGFVAGLRE 162
Qy 169 EVVYTSMAKWIYTKGSKRHYGTQKQNRSSRSHITFRMILESREK----GEPSNCEGS 224
Db 163 EIVASPPQVLEMEFGEHSHRIGETNMNLYSRSHITFRMILESREKQKQMDGEGVNSCD-A 221
Qy 225 VKVSHNLVDLAGERRAQTGAAGVRLKEGNCINRSFLFGVVKLSDG--QVGGFTNY 282
Db 222 VRVSVNLVDLAGERRAQTGAAGVRLKEGSHINKSLMTLGTIVIKKSEGVEQTGGHPY 281
Qy 283 RDSKLTILQNSLGNPKTRITICIT--PVSFDETLTALQFASAKYMKNTPYVNEVSTD 340
Db 282 RDSKLTILQALGNANTALICNTLAPIHADETKSLQFASRALRVNCAHVNEILTD 341
Qy 341 EALKRYRKEIMDLKKQL-----EVSLETR-----AQAMEKDLAQLEEKDLLO-- 386
Db 342 AALLKQKQKETEELRSKLTSHSDHSEIILNRLNTLKLSELERIALELEEEKKAQAQ 401
Qy 387 -----KVQNEKIENLRLMTVSS 404
Db 402 RERVLOEQAKKIKNLSSMVLISN 424
RESULT 6
ID Q9LQ62 PRELIMINARY; PRT; 888 AA.
AC Q9LQ62;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE T30R16.9.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,
RA Ronning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,
RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RA "Arabidopsis thaliana chromosome III BAC F14P13 genomic sequence.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC009400; AA02823.1; -;
DR HSP; P33176; 18G2.
DR GO; GO:0005871; C:kinesin complex; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR GO; GO:0007017; P:microtubule-based process; IEA.
DR InterPro; IPR001752; kinesin_motor.
DR Pfam; PF00225; Kinesin; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
DR PROSITE; PS00067; KINESIN MOTOR DOMAIN2; 1.
SQ SEQUENCE 459 AA; 51872 MW; 768487E9B7038E6A CRC64;
Query Match 34.0%; Score 844.5; DB 10; Length 888;
Best Local Similarity 40.0%; Pred. No. 1.1e-43;
Matches 220; Conservative 70; Mismatches 169; Indels 91; Gaps 13;
Qy 6 VAVCVVRPLNSREESIGETAQVYWKYTDNNVI-----Y 38
Db 4 ICVAVVRP-----PAPENGASLWKVEDNRISLHSLDTPITTASHAFVSGISISTDLI 57
Qy 39 QVDSKSNF-----DRVPHGNETTKVVEETAAPIIDSAIQYNGTIFAYGCTA 88
Db 58 EIVSLFLFSGVYVFFLLFADHVFDESSNASVVELTKDIIHAAVEGFGTAFAYGTS 117
Qy 89 SKGTYTWMGSDHLGVIPRAIHDFQKIKFPDFREFLLRVSYMEIYNETITDLCCGTQK 148
Db 118 SKGTYTWMGSDHLGVIPRAIHDFQKIKFPDFREFLLRVSYMEIYNEEINDLL--AVEN 175
Qy 149 KPLIREDVNRNVYADLTEEVYTSMAKWIYTKGSKRHYGTQKQNRSSRSHITFRM 208
Db 176 QRLQTHEHLRGVVFAGLKEEIVSDAEQILKIDSGEVNRHFGTNNMVHSSRSHITFRM 235
Qy 209 -----ILESREKGEPSNCEGSVKVSHNLVDLAGERRAQTGAAGVRLKEGNC 256
Db 216 VRFSYERDILLVIESR--GKDNSSDAIRVSVNLVDLAGERSAKTAKGAGVRLQEGKY 293
Qy 257 INRSILFIQGVVKLSDG--QVGGFTNYRDSKLTILQNSLGNPKTRITICTTPVS--FD 313
Db 294 INKSMILGNVINKLSDSTKLRAHPIYRDSKLTILQALGNNAKTCIITIAPEHHIE 353

```

QY 314 ETLTALQFASTAKYKNTPTVYNEVSTDEALKRYRKEIMDLKKOLEEVSLETRAQ----- 368
Db 354 ESKTLOFASRAKRITCAQVNEILITDALLKROKLEIEELRMLKLGSHAVLQOEILNL 413
QY 369 -----AMEKDQLAQLLEEDLQKQVNE-----KIENLTRMLVTSSSLTLQOELKA 414
Db 414 SNQMLKVELECEERLKTOLLEEKRKQKEQNCIKQEQMKIENLNN-FVNSDF----- 464
QY 415 KRKRVTWCLGKINKMKNNSVADQFNPTNTITTKTKLSINMLREIDSVCSGESDVSFNT 474
Db 465 KRQNSDFIISRKTPDGLCNVNDTSDVGTCPFKSASRFSVVARSNNYSGLSDFSPMVHS 524
QY 475 L-DTLSELEW 483
Db 525 LGDVAEDTW 534

RESULT 7
ID Q869B8 PRELIMINARY; PRT; 1885 AA.
AC Q869B8;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Kinesin-related protein K4.
GN K4.
OS Dictyostelium discoideum (slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98359834; PubMed=9693369;
RA de Hostos E.L., McCaffrey G., Sugang R., Pierce D.W., Vale R.D.;
RT "A developmentally regulated kinesin-related motor proteins from
RT Dictyostelium discoideum."
RL Mol. Biol. Cell 9:2093-2106(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Suyama E., Sutoh K.;
RT "Kinesin-related proteins from Dictyostelium."
RT Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB102780; BAC56912.1; -
DR GO; GO:0036020; C:membrane; IEA.
DR GO; GO:0005875; C:microtubule associated complex; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR InterPro; IPR001752; F:motor activity; IEA.
DR InterPro; IPR003345; M_repeat.
DR Pfam; PF00225; kinesin; 1.
DR Pfam; PF02370; M; 6.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR PROSITE; PS00667; KINESIN_MOTOR_DOMAIN2; 1.
SQ SEQUENCE 1885 AA; 218216 MW; 15686AED3B007EC7 CRC64;

Query Match 32.3%; Score 801; DB 5; Length 1885;
Best Local Similarity 39.7%; Pred. No. 1.4e-40;
Matches 193; Conservative 79; Mismatches 144; Indels 70; Gaps 9;

QY 2 EBGAVACVRVPLNSRESISLGETAQVYWKTDNNVI-YQVDGSKSFNDRVFGNETTKN 60
Db 19 ELNKIKVAIRVPLNSRE--LGIDQKIPWSISKDTLSLQNPNNINFTYDVFQIDSNITD 76
QY 61 VVEETAAPIDGALQNGTIFAYGQTASGKYTWGMSDHLGVIPRAIHDFOKIK-KF 119
Db 77 VYNAIKLSVNSLNGINGTIFAYGQTSKGKFSMGRTSIPGIKLSIKDFKSIEDSI 136
QY 120 PPREPLRVSYMEIYNETITDLGGT-QQKQPLIINEDVRNRYVADLTPEEVVYTSMAL 178
Db 137 LEKDYLKVSYLEIYNEEIKOLLNPTISNKKLIKIHEDIVKGVVAVNLKEEIVISPDQIF 196

```

```

QY 179 KWIITGKSRHYGTGMNQRSSRSHITFRMILESRKEGPEPCNCEGSKVXVSHNLNVLDA 238
Db 197 ALMNFGEERRHIGSTWMDSSSRSHITFRWQIQSTCKQ-----NGTIQMSLTIVDLA 251
QY 239 ERAAQTGAAGVRLKEGNCINRSFILQVYIKLSDGGVGGFINYRDSKLTILQNSLGGN 298
Db 252 ERVSTGAEGVRLKEGTHINKSLMTLSKVISKLESEKTTQHPYRDSKLTILQPSLGGN 311
QY 299 PKRIITCTTPVSF--DETLTALQFASTAKYKNTPTVYNEVSTDEALKRYRKEIMDLKK 356
Db 312 SKTAILCTITPATTHQEEISITLQFAKRAKVKNTYKINQVADANTMLKKYSEISELELQ 371
QY 357 QLEEVSLLETRAQAMEKDQLAQLLEEKDOLLQKVQEKIENLTRMLVTSSSL----- 406
Db 372 QL-----VKSEINSLLRNTISTQEISSNNFKLGWK 402
QY 407 -----TLQELKAKRKRRVTW-----CLGKINKMKNNSVADQFNPTNTIT 447
Db 403 RFNDATIGGSLINENKKKKRNTLDPFYLKDKIIKKIRKSENQIKKIKNSENNIS 462
QY 448 KTHKLS 453
Db 463 SSSNS 468

RESULT 8
QYVK10 PRELIMINARY; PRT; 2013 AA.
ID QYVK10
AC QYVK10;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Cmet protein
GN CMET OR C6392.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballou R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng X., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

```

[illegible]

Dbb : : : : :
 449 NILGSLD--IGTSENIS 464

RESULT 10
Q8RWW4
ID Q8RWW4 PRELIMINARY; PRT; 1055 AA.
AC Q8RWW4;
DT 01-JUN-2002 (TEMBLrel. 21, Created)
DT 01-JUN-2002 (TEMBLrel. 21, Last sequence update)


```

QY 301 TRIICITTPV--SFDETLTALQPASTAKYMKNTPTVYNEVSTDBALLKRYRKEIMDLKKQL 358
Db 386 VSLICITTPASSSEETHNLTKASRAKSIIEIYASRNQIIDKSLIKKYQREISLKL 445
QY 359 EVV-----SLTRAQAMEKQ--LAQLLEFKDLQKQVNEKINLRLMLVTSS 405
Db 446 DQLRRGLMVGVSHEELMSLQQLBEGQVKQSRLEEEBAKAALMSRIQKLTLLIVST- 504
QY 406 LTLQOELKAKRKRVTWCLGKINKMKNSYADQFNIPNTITTKTKLSINLLREID-ESV 464
Db 505 -----KNSIPGSGDIP-----THQRSAGSKDDKFDLSL 533
QY 465 CSBSDFV---SNTLDTLSE 480
Db 534 LLESNDNLGSPSSTLALLSE 552

RESULT 12
P87198 PRELIMINARY; PRT; 1459 AA.
AC P87198;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Kinesin motor protein.
GN KIN1.
OS Ustilago maydis (Smut fungus).
OC Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;
OC Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.
CX NCBI_TaxID=5270;
[1]
SEQUENCE FROM N.A.
MEDLINE=97361828; PubMed=9218789;
RA Lehmler C., Steinberg G., Snetelaar K.M., Schliwa M., Kahmann R.,
RA Bolker M.;
RA "Identification of a motor protein required for filamentous growth in
RA Ustilago maydis.";
RL EMBO J. 16:3464-3473(1997).
DR EMBL; U92844; AAB63336.1; -.
DR PIR; T30196; T30196.
DR HSP; P17119; 3KAR.
DR GO; GO:0005871; C:kinesin complex; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR GO; GO:0007017; P:microtubule-based process; IEA.
DR InterPro; IPR001752; kinesin_motor.
DR Pfam; PF00225; Kinesin; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR PROSITE; PS0067; KINESIN_MOTOR_DOMAIN2; 1.
SQ SEQUENCE 1459 AA; 156932 MW; 281874B0100F5CD0 CRC64;

Query Match 30.6%; Score 759; DB 3; Length 1459;
Best Local Similarity 38.2%; Pred. No. 3.8e-38;
Matches 191; Conservative 71; Mismatches 128; Indels 110; Gaps 14;

QY 6 VAVCVVRPLNSREESLGETAQVYKTDN--NVLYQVD----- 41
Db 237 VVVCVRMRP--SRASSDSSEASV--WNCDSKRNRPFTTEHHPAIAKRTTSSERAGAGASIA 293
QY 42 -----GSKSFNDFRVFHNGETTKNVEEIAAIPIDSAIQNGYNTIFAYGQ 86
Db 294 AAPSSHDLHDHEDPTSSYHFQDKLITGAQITDDMYHSHIAPVVEAAVEGYNGTVFAYGQ 353
QY 87 TASGTYTNMGSEDLHGVIPRAIHDI FOKIKFPDRFLLRVSYMEIYNEITDLLCGTQ 146
Db 354 TGSQKTHMTSGSDAEPGVPRAVEQIFQMIKDEPDRELLRVSYLEIYNETLKKLLA--- 410
QY 147 KMKP-----IREDVNRNVVADLTSEEV 171
Db 411 PLPLITGTGSSLQTTDRPASPIKGGSSHAAGQSQTILRIEDQKSRVITGLREIV 470

```

```

QY 172 YTEAMALKWTTKGEKSRHRYGETKONORSSRSHTIFRMILESREKGPSCGSKVSHLN 231
Db 471 TDANTVLCILQGDHVRHVGATDWNERSRSRSHCVFQLTIESRPAFSAKSE--VRISQLN 528
QY 232 LVTLASERAAQTGAAGVRLKEGCNTNRSILFILGOVIKKLSLDGQVG--FINYRDSKLTR 289
Db 529 LIDLASERAA---SQAERKEGAFINKSLTLTGTVIKUTEPEVNGDAHIIPRDSKLTR 585
QY 290 ILQNSLGCNPKRIICITITPVS--FDETLTALQPASTAKYMKNTPTVYNEVSTDEALKRY 347
Db 586 ILTSLSGNARIAVICTLSPDTEHANETLSLTFKGEKCLVVTAKKGTAMDDKALLQKY 645
QY 348 RKEIMDLKKOLEE-----VSLTRAQAMEK-DQLAQLE-----EKDLQKQV 389
Db 646 RKELDALRAKLEANGPSPNEVMTIVSABASKESQQLDLQNAQOEAAQREVEDMQKR 705
QY 390 NE---KLENTRLMLVTSSSL 406
Db 706 SHLKAQIEHTRLLTLTSQSV 725

RESULT 13
Q9SJUU PRELIMINARY; PRT; 1058 AA.
AC Q9SJUU; Q94BQ1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative kinesin heavy chain.
GN AT2G21380.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
CX NCBI_TaxID=3702;
[1]
SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Lin X., Kaul S., Shea T.P., Fujii C.Y., Shen M., VanAken S.E.,
RA Barnstead M.E., Mason T.M., Bowman C.L., Roming C.M., Benito M.-I.,
RA Carrera A.J., Creasy T.H., Buell C.R., Town C.D., Nierman W.C.,
RA Fraser C.M., Venter J.C.;
RA Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
[2]
SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA Dale J.M., Gibson H.A., Goldsmith A.D., Jiang P.X., Lee J.M.,
RA Quach H.L., Tang C., Toriumi M., Yu G., Bowser L., Carninci P.,
RA Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T., Kamiya A.,
RA Karlin-Neumann G., Kawai J., Kim C., Koesema E., Lam B., Lin J.,
RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Tracy S.E.,
RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RA "Full length cDNA of gene F3K23.14/At2g21380 (GI:4567271).";
RA Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
[3]
SEQUENCE FROM N.A.
RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Hsuan V.W., Lee J.M.,
RA Quach H.L., Tang C., Toriumi M., Wallender E.K., Wong C., Wu H.C.,
RA Yu G., Yuan S., Carninci P., Chen H., Cheuk R., Hayashizaki Y.,
RA Ishida J., Jones T., Kamiya A., Kawai J., Kim C.J., Narusaka M.,
RA Nguyen M., Palm C.J., Sakurai T., Sacou M., Seki M., Shinn P.,
RA Southwick A., Tripp M.G., Wu T., Shinozaki K., Davis R.W., Ecker J.R.,
RA Theologis A.;
RA "Arabidopsis Open Reading Frame (ORF) Clones.";
RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006841; AAD23684.2; -.
DR EMBL; AY039966; AAK64143.1; -.
DR PIR; E84600; E84600.
DR HSP; P17119; 3KAR.
DR GO; GO:0005871; C:kinesin complex; IEA.

```


QY	5	AVAVCVVRPLNSRRESIGETAQVYWKTDNNVIQV----	DG-SKSFNFRVPHGNETT	58
DB	4	SVRVAVRCRPFNOREKDLNTLVCV-GMPFN--VGQVNLNAPDGAADFTFGAYPMDSTG	60	
QY	59	KNVVEELAAPIIDSAIQGVNGTIFAYGOTASGKTYTMMGSED----	HLGVIPIRAIHDFQK	115
DB	61	EQIYNDIVFVPLVENVIEGYNGTVFAYGOTGSKTFSMOGLIETIPAQRGVI	PRAFDHIFTA	120
QY	116	TKKFPDRPFLRVSYMEIYNETIDLLCGTQKQKPLIITREDVNRNVYVADLT	TEEVVYTS	175
DB	121	TATTENVKFLVHCYSLEYIENVEVRDLGADNKKQ-LEIKQPDGRGVYAGLSMHVCHDVP	179	
QY	176	MALKWITGKSRHYGETKMNORSRSHITIFMILESREKGFSPNCEGSKVSHNLVDL	235	
DB	180	ACKELMTRGFNNRHVGATIMNKKDSRSHSIFVYVEGMT-----	TGSIIRMGKUNLVDL	233
QY	236	AGSERRAOTGAAGVRLKEGCNINRSLFILGVIKKLSQGVGFINYRDSKLTRIQNSL	295	
DB	234	AGSERQSKTGATGDRLKEATKINLSLSALGVISALVDGK-SKHIPYRDSKLTRIQDSL	292	
QY	296	GGNPKTRITCTITPVS--FDETLTALQFASTAKYMKNTPTVYNEVSTDEALLKRYRKEIMD	353	
DB	293	GGNTKTIMIACVSPSSDNYDETLSLTVRANRAKNIKNPETINEDPKD-ALLREYOEIAR	351	
QY	354	LKKQLEEVSLFTRAQAM-----EKDQLAQLER--KDLLQKVONEKIENLTRLMLVTSS	404	
DB	352	LKSMYQPGAVGVGAQPAQDAFSTEEERKKLUREFEEMDLRGEYEREQ-----	ITS	401
QY	405	SLTUQELKXKXK--RRVTVCLGKIN----KMKNSNYADQF-----	NIP	442
DB	402	KAEQXQLESRLADYERANANLDNLNPEEAKKIQQLQDQFIFGEEAGNTQLKQKEMQL	461	
QY	443	TNITTKTKHLSINLREIDESVCSGSDVFSNTLDLTLSEI	481	
DB	462	KBAETKTOKLAALNKHVDPL--QVYSTTQEKLDV	497	

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 09:26:01 ; Search time 20.4051 Seconds

(without alignments)
4694.096 Million cell updates/sec

Title: US-10-045-631B-88_COPY_2_340

Perfect score: 1744

Sequence: 1 AEEGAVAVCVVRPLNSREE.....QFASTAKYKNTFYNEVNST 339

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_23Jan04:*

- 1: Geneseqp1980s:*
- 2: Geneseqp1990s:*
- 3: Geneseqp2000s:*
- 4: Geneseqp2001s:*
- 5: Geneseqp2002s:*
- 6: Geneseqp2003as:*
- 7: Geneseqp2003bs:*
- 8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1744	100.0	2633	4	ABG06505	Abg06505 Novel hum
2	1744	100.0	2663	4	AAm39097	AAm39097 Human pol
3	1635.5	93.8	2688	4	AAm40883	AAm40883 Human pol
4	1299.5	74.5	2954	2	AAy01632	AAy01632 Amino aci
5	781	44.8	348	4	ABU53208	ABU53208 Human cel
6	781	44.8	366	4	ABU53125	ABU53125 Intracell
7	707	40.5	2013	4	ABG70991	ABG70991 Human tar
8	668	38.3	473	5	ABG70992	ABG70992 Human tar
9	668	38.3	522	5	ABG70991	ABG70991 Human tar
10	668	38.3	1232	7	ADD49338	ADD49338 Human lun
11	668	38.3	1232	7	ADD18924	ADD18924 Human lun
12	658	37.7	1232	7	ADD49937	ADD49937 Human lun
13	658	37.7	1232	7	ADD49932	ADD49932 Human lun
14	652	37.4	1232	5	ABG70990	ABG70990 Human tar
15	651	37.3	1234	5	ABG70993	ABG70993 Human tar
16	649.5	37.2	1029	5	AAE17786	AAE17786 Human kin
17	649.5	37.2	1029	7	ADC10190	ADC10190 Human MOV
18	646.5	37.1	677	4	ABG65183	ABG65183 Drosophil
19	641.5	36.8	1038	5	AAm48337	AAm48337 Human kin
20	621.5	35.6	1034	3	AAQ31112	AAQ31112 Arabidops
21	621.5	35.6	1069	3	AAQ31111	AAQ31111 Arabidops
22	621.5	35.6	1121	3	AAQ31110	AAQ31110 Arabidops
23	620.5	35.6	329	5	AAE17787	AAE17787 Human kin
24	618	35.4	784	2	AAy06618	AAy06618 Thermomyc
25	615.5	35.3	411	2	AAW72745	AAW72745 Drosophil

ALIGNMENTS

RESULT 1

ABG06505
ID ABG06505 standard; protein; 2633 AA.

XX AC ABG06505;

XX DT 13-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #6496.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX WO200175067-A2.

XX PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US008631.

PR 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.

XX (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

DR N-PSDB; AAS70692.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.

XX Claim 20; SEQ ID NO 36864; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging

Aaw72744 Drosophil
Aaw72746 Drosophil
Abb63485 Drosophil
Abb67088 Kinesin h
Aae14402 Human HsK
Aau79592 Human HsK
Aae14400 Human kin
Aau79590 Human kin
Aab48222 Human bla
Adb80468 Ovarian c
Adc35116 Human bre
Ade55349 Rat Prote
Adb67093 Kinesin h
Abb62962 Drosophil
Abb59245 Drosophil
Aag67418 Amino aci
Abg72693 Fruitfly
Adb67091 Kinesin h
Aaul9569 Human dia
Abp51294 Human MDD

26 615.5 35.3 441 2 AAW72744
27 615.5 35.3 975 2 AAW72746
28 615.5 35.3 975 4 ABB63485
29 615.5 35.3 975 7 ADB67088
30 615 35.3 409 5 AAE14402
31 615 35.3 409 5 AAU79592
32 615 35.3 1388 5 AAE14400
33 615 35.3 1388 5 AAU79590
34 615 35.3 1388 6 ABR48222
35 615 35.3 1388 7 ADB80468
36 615 35.3 1388 7 ADC35116
37 613 35.1 796 7 ADE55349
38 612.5 35.1 1031 7 ADB67093
39 609 34.9 1921 4 ABB62962
40 605 34.7 1048 4 ABB59245
41 605 34.7 1066 4 AAG67418
42 605 34.7 1066 6 ABG72693
43 604 34.6 963 7 ADB67091
44 603.5 34.6 757 4 AAU19569
45 603.5 34.6 757 5 ABP51294

of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences)

QY	1	2	61	62	121	122	181	182	241	242	301	302
Query Match	100.0%	100.0%	Score 1744; DB 4; Length 2633;									
Best Local Similarity	100.0%	Pred. No. 2.2e-161;										
Matches 339; Conservative 0; Mismatches 0; Indels 0; Gaps 0												
QY	1	2	61	62	121	122	181	182	241	242	301	302
Db	1	2	61	62	121	122	181	182	241	242	301	302
QY	1	2	61	62	121	122	181	182	241	242	301	302
Db	1	2	61	62	121	122	181	182	241	242	301	302
QY	1	2	61	62	121	122	181	182	241	242	301	302
Db	1	2	61	62	121	122	181	182	241	242	301	302
QY	1	2	61	62	121	122	181	182	241	242	301	302
Db	1	2	61	62	121	122	181	182	241	242	301	302
QY	1	2	61	62	121	122	181	182	241	242	301	302
Db	1	2	61	62	121	122	181	182	241	242	301	302
QY	1	2	61	62	121	122	181	182	241	242	301	302
Db	1	2	61	62	121	122	181	182	241	242	301	302
QY	1	2	61	62	121	122	181	182	241	242	301	302
Db	1	2	61	62	121	122	181	182	241	242	301	302
QY	1	2	61	62	121	122	181	182	241	242	301	302
Db	1	2	61	62	121	122	181	182	241	242	301	302
QY	1	2	61	62	121	122	181	182	241	242	301	302
Db	1	2	61	62	121	122	181	182	241	242	301	302
QY	1	2	61	62	121	122	181	182	241	242	301	302
Db	1	2	61	62	121	122	181	182	241	242	301	302
QY	1	2	61	62	121	122	181	182	241	242	301	302
Db	1	2	61	62	121	122	181	182	241	242	301	302
QY	1	2	61	62	121	122	181	182	241	242	301	302
Db	1	2	61	62	121	122	181	182	241	242	301	302
QY	1	2	61	62	121	122	181	182	241	242	301	302
Db	1	2	61	62	121	122	181	182	241	242	301	302
QY	1	2	61	62	121	122	181	182	241	242	301	302
Db	1	2	61	62	121	122	181	182	241	242	301	302
QY	1	2	61	62	121	122	181	182	241	242	301	302
Db	1	2	61	62	121	122	181	182	241	242	301	302
QY	1	2	61	62	121	122	181	182	241	242	301	302
Db	1	2	61	62	121							

RESULT 2
AAM39097
ID AAM39097 standard; protein; 2663 AA.
XX
AC AAM39097;

XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 2242.

Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
peripheral nervous system; neuropathy; central nervous system; CNS;
Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
anyother's lateral sclerosis; Shy-Drager Syndrome; chemotactic;
chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
leukaemia.

XX
OS Homo sapiens.

AA
PN
WO200153312-A1.

26-JUL-2001.

XX
PF
26-DEC-2000: 2000WO-US034263.

XX
PR 23-DEC-1999: 99US-00471275.

PR 21-JAN-2000; 2000US-00488723;
PR 25-APR-2000; 2000US-00552317;

PR 20-JUN-2000; 2000US-0035804Z;
PR 19-JUL-2000; 2000US-0062031Z.

Pf Pf Pf Pf Pf X X Pf X Pf Pf Pf X D D X P P X X C C C C C C C C C C C C C X S

03-AUG-2000; 2000US-00653450.
14-SEP-2000; 2000US-00562191.
19-OCT-2000; 2000US-00593036.
29-NOV-2000; 2000US-00727344.

(HYSE-) HYSEQ INC.

Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren P, Wang D;
Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
Zhou P, Goodrich R, Drmanac RT;

WPI; 2001-442253/47.
N-PSDB; AAI58253.

Novel nucleic acids and polypeptides, useful for treating disorders such as central nervous system injuries.

Example 4; SEQ ID NO 2242; 10078pp; English.

The invention relates to human nucleic acids (AAI57798-AAI61369) and the encoded polypeptides (AAM38642-AAM42213) with nootropic, immunosuppressant and cytostatic activity. The polynucleotides are useful in gene therapy. A composition containing a polypeptide or polynucleotide of the invention may be used to treat diseases of the peripheral nervous system, such as peripheral nervous injuries, peripheral neuropathy and localised neuropathies and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager Syndrome. Other uses include the utilisation of the activities such as: Immune system suppression, Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, cancer diagnosis and therapy, drug screening, assays for receptor activity, arthritis and inflammation, leukaemias and C.N.S disorders. Note: The sequence data for this patent did not form part of the printed specification

Sequence 2563 AA;

```
Query Match      100.0%; Score 1744; DB 4; Length 2663;
Best Local Similarity 100.0%; Pred. No. 2.3e-161;
Matches 339; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy	1	A	E	G	A	V	A	V	C	V	R	P	L	N	S	R	E	S	I	G	E	T	A	O	V	N	K	T	D	N	N	V	I	Q	V	D	G	S	K	S	F	N	F	D	R	V	F	H	C	N	E	T	K	N	60		
Db	2	A	E	G	A	V	A	V	C	V	R	P	L	N	S	R	E	S	L	E	T	A	O	V	N	K	T	D	N	N	V	I	Q	V	D	G	S	K	S	F	N	F	D	R	V	F	H	C	N	E	T	K	N	61			
Qy	61	V	Y	E	E	I	A	A	P	I	D	S	A	I	Q	Y	N	G	T	I	F	A	Y	G	T	A	S	K	T	T	M	G	S	E	D	H	L	G	V	P	I	P	A	I	H	I	F	O	K	I	K	F	P	120			
Db	62	V	Y	E	E	I	A	A	P	I	D	S	A	I	Q	Y	N	G	T	I	F	A	Y	G	T	A	S	K	T	T	M	G	S	E	D	H	L	G	V	P	I	P	A	I	H	I	F	O	K	I	K	F	P	121			
Qy	121	D	R	E	F	L	L	R	V	S	W	E	I	N	E	T	I	T	O	L	L	C	G	T	O	K	K	P	L	I	I	R	E	D	V	N	R	N	V	V	A	D	L	T	E	E	V	V	T	S	E	M	A	L	K	W	180
Db	122	D	R	E	F	L	L	R	V	S	W	E	I	N	E	T	I	T	O	L	L	C	G	T	O	K	K	P	L	I	I	R	E	D	V	N	R	N	V	V	A	D	L	T	E	E	V	V	T	S	E	M	A	L	K	W	181
Qy	181	I	T	K	E	S	R	H	Y	G	E	T	K	M	O	N	S	R	S	H	T	I	F	M	L	E	S	R	E	K	G	P	S	C	E	G	S	V	K	V	K	Y	S	H	L	N	L	V	D	L	A	G	S	R	240		
Db	182	I	T	K	E	S	R	H	Y	G	E	T	K	M	O	N	S	R	S	H	T	I	F	M	L	E	S	R	E	K	G	P	S	C	E	G	S	V	K	V	K	Y	S	H	L	N	L	V	D	L	A	G	S	R	241		
Qy	241	A	A	O	T	G	A	A	V	L	K	E	C	N	I	N	S	I	F	I	L	G	O	V	I	K	L	S	D	O	V	G	F	N	I	R	D	S	K	L	T	R	I	L	O	N	S	L	G	N	P	K	300				
Db	242	A	A	O	T	G	A	A	V	L	K	E	C	N	I	N	S	I	F	I	L	G	O	V	I	K	L	S	D	O	V	G	F	N	I	R	D	S	K	L	T	R	I	L	O	N	S	L	G	N	P	K	301				
Qy	301	T	R	I	C	T	T	P	S	E	D	E	T	I	T	A	L	O	F	A	S	T	A	K	Y	M	O	N	T	P	P	Y	N	E	V	S	T	339																			
Db	302	T	R	I	C	T	T	P	S	E	D	E	T	I	T	A	L	O	F	A	S	T	A	K	Y	M	O	N	T	P	P	Y	N	E	V	S	T	340																			

RESULT 3

AAM40883

ID AAM40883 standard; protein; 2688 AA.

AC AAM40883;

DT 22-OCT-2001 (first entry)

Human polypeptide SEQ ID NO 5814.

Human; neotrophic; immunosuppressant; cytostatic; gene therapy; cancer; peripheral nervous system; neuropathy; central nervous system; CNS; Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic; chemokinetic; thrombolytic; drug screening; arthritis; inflammation; leukaemia.

OS Homo sapiens.

XX NC0200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US034263.

XX 23-DEC-1999; 99US-00471275.

PR 21-JAN-2000; 2000US-00488725.

PR 25-APR-2000; 2000US-00552317.

PR 20-JUN-2000; 2000US-00598042.

PR 19-JUL-2000; 2000US-00620312.

PR 03-AUG-2000; 2000US-00653450.

PR 14-SEP-2000; 2000US-00662191.

PR 19-OCT-2000; 2000US-00693036.

PR 29-NOV-2000; 2000US-00727344.

XX (HYSE-) HYSEQ INC.

PA Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D; Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA; Zhou P, Goodrich R, Drmanac RT;

PI WPI: 2001-442253/47.

DR N-PSDB; AA160039.

XX Novel nucleic acids and polypeptides, useful for treating disorders such as central nervous system injuries.

PT Example 2; SEQ ID NO 5814; 10078pp; English.

PS The invention relates to human nucleic acids (AA157798-AA161369) and the encoded polypeptides (AA138642-AA142213) with neotrophic, immunosuppressant and cytostatic activity. The polynucleotides are useful in gene therapy. A composition containing a polypeptide or polynucleotide of the invention may be used to treat diseases of the peripheral nervous system, such as peripheral nervous injuries, peripheral neuropathy and localised neuropathies and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager Syndrome. Other uses include the utilisation of the activities such as: Immune system suppression, Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, cancer diagnosis and therapy, drug screening, assays for receptor activity, arthritis and inflammation, leukaemias and C.N.S disorders. Note: The sequence data for this patent did not form part of the printed specification

XX Sequence 2688 AA;

Query Match 93.8%; Score 1635.5; DB 4; Length 2688;

Best Local Similarity 95.3%; Pred. No. 1.1e-150;

Matches 326; Conservative 2; Mismatches 11; Indels 3; Gaps 3;

QY 1 AEEGAVACVVRPLNSRESLGETAQYWKTDNRVIVQVDSKSFNDRVFGHNETYKN 60

DB 23 AEEGAVACVVRPLNSRESLGETAQYWKTDNRVIVQVDSKSFNDRVFGHNETYKN 82

QY 61 VVEETAAPIIDSAIOGNGTTFAYGQTASGKTYTMMGSEDLGVTPRA-INDIF-QKIKK 118

DB 83 VVEETAAPIIDSAIOGNGTTFAYGQTASGKTYTMMGSEDLGVTPRA-INDIF-QKIKK 142

QY 119 -FPDREFLLRVSYMEIYNETITDLCTGQKMKPLIREDVNRNVYADLTVEEVYTSEMA 177

DB 143 VFIDREFLLRVSYMEIYNETITDLCTGQKMKPLIREDVNRNVYADLTVEEVYTSEMA 202

QY 178 LKWITKGEKSRHYGETKMQRSRSHITFRMILESREKGFSPNCEGSKVSHLNLVDLAG 237

DB 203 LKWITKGEKSRHYGETKMQRSRSHITFRMILESREKGFSPNCEGSKVSHLNLVDLAG 262

QY 238 SERAATGAGVRLKEGCMNRSFLFGLQVKKLSDGQVGGFNYRDSKUTRILQNSLGG 297

DB 263 SERAATGAGVRLKEGCMNRSFLFGLQVKKLSDGQVGGFNYRDSKUTRILQNSLGG 322

QY 298 NPKTRIICITPVSFDETLTALQFASTAKYMKNTPVYNEVST 339

DB 323 NPKTRIICITPVSFDETLTALQFASTAKYMKNTPVYNEVST 364

RESULT 4

AA01632

ID AA01632 standard; protein; 2954 AA.

AC AA01632;

XX 22-JUN-1999 (first entry)

XX Amino acid sequence of centromere-associated protein-E (CENP-E).

XX CENP-E; centromere-associated protein-E; ATPase activity; plus end-directed microtubule motor activity; chromosome congression; microtubule binding activity; chromosome movement; mitosis; cell proliferation; tumor; metastasis; vascular malfunction; inflammatory disease; immune disease; angiogenesis; hypertension; restenosis; fungal infection; selective herbicide; fungicide; insecticide; plant growth regulator; activator; cancer cell marker.

XX Xenopus sp.

XX WO9913061-A1.

XX 18-MAR-1999.

XX 10-SEP-1998; 98WO-US019231.

XX 11-SEP-1997; 97US-0058645P.

XX (REGC) UNIV CALIFORNIA.

XX Wood KW, Sakowicz R, Goldstein LSB, Cleveland DW; WPI: 1999-229233/19.

XX N-PSDB; AA26819.

XX Centromere-associated protein-E and related nucleic acid.

XX Claim 5; Page 66-67; 77pp; English.

XX The present sequence represents CENP-E (centromere-associated protein-E) of Xenopus. The protein has at least one of plus end-directed microtubule motor activity, ATPase (adenosine triphosphatase) activity and microtubule binding activity. CENP-E is the motor that powers chromosome movement toward microtubule plus ends and is essential for congression of chromosomes during mitosis. Modulators of CENP-E can thus control cell proliferation. Agents that modulate CENP-E activity are lead therapeutic, bioagricultural and diagnostic agents, e.g. for treatment of unwanted cell proliferation (typical of many examples are tumors and metastases; vascular malfunction; inflammatory and immune diseases; angiodysplasia; hyperextension; restenosis; and fungal infections), also as plant-protection agents (selective herbicides, fungicides and insecticides) and plant growth regulators or activators for improving yields. CENP-E is also a diagnostic marker for dividing cells, including cancer cells

XX Sequence 2954 AA;

Query Match 74.5%; Score 1299.5; DB 2; Length 2954;

Best Local Similarity 73.5%; Pred. No. 1.5e-117;
Matches 250; Conservative 42; Mismatches 43; Indels 5; Gaps 2;
QY 1 AEGAVAVCVVRPLNSRESLSGETAQVYKTDNNVYQVDSKSFNDRVFGHNETTKN 60
Db 2 SEGDAVKVCVRVPLIQREQ--GDQANLOWKAGNNTISQVDTGKSFNDRVFNESHSTSQ 59
QY 61 VYREIAPIIDSALQGVNGTIFAYGQTASGKTYTMMGSEDLHGVIPRAIHDIFOKIKKFP 120
Db 60 IYQEIAPVIRSAIQNGTIFAYGQTSSGKTYTMMGTSPNSLGIIPQAIQEVFKIQEIP 119
QY 121 DREFLLRVSYMEIYNETITDLCCTQKPKPIIREDVARNVYVADLVEEVVYTSSEALKW 180
Db 120 NREFLRVSYMEIYNETVOLLCDRRKKPLEIREDEFNRNVYVADLVEELVMVPEHVIOW 179
QY 181 ITKGEKSRHGETKQNRSSRSHTIFRMILESREKGEPS--NCEGSVKYSHLNLVDLAG 237
Db 180 IKKEKNRHYGETKMDHSSRSHTIFRMIVESRDNDPTNSGCDGAVWWSHLNLVDLAG 239
QY 238 SERAAQTGAAGVRLKEGCNINRSLFILQGVIKKLSGQGVGFVINYRDSKLTILONSLGG 297
Db 240 SERASQTGAEGVRLKEGCNINRSLFILQGVIKKLSGQAGGFVINYRDSKLTILONSLGG 299
QY 298 NPKTRIICTIPVDFEFLTALQFASAKYMKNTPYNEV 337
Db 300 NAKTVIICTIPVDFEFLTALQFASAKYMKNTPYNEV 339

RESULT 5
ABU53208
ID ABU53208 standard; protein; 348 AA.
XX AC ABU53208;
XX DT 14-APR-2003 (first entry)
XX DE Human cell cycle-associated DKFZphtes3_35b4 homologue #7.
XX KW Human; gene therapy; vaccine; disease treatment; detection.
XX OS Homo sapiens.
XX PN WO200112659-A2.
XX PD 22-FEB-2001.
XX PF 18-AUG-2000; 2000WO-IB001496.
XX PR 18-AUG-1999; 99US-0149499P.
XX PR 28-SEP-1999; 99US-0156503P.
XX PA (GEHU-) GERMAN HUMAN GENOME PROJECT.
XX PI Wiemann S;
XX DR WPI; 2001-327840/34.
XX PT Nucleic acids having the sequences of clones isolated from libraries of different human tissues, useful in recombinant DNA methodologies.
XX PS Example III; Page 829; 1095pp; English.
XX PS This invention describes novel polynucleotides and polypeptides isolated from human cDNA libraries which can be used for gene therapy or in vaccines. The polynucleotides of the invention and antibodies encoded by them may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate polypeptide expression. The products of the invention may also be used to identify modulators of expression and activity and to down regulate expression and activity. The antibodies of the invention may also be used as diagnostic agents for detecting the presence of polypeptides in samples. This sequence represents a homologue of a polypeptide described in the disclosure of the invention

SQ Sequence 348 AA;
Query Match 44.8%; Score 781; DB 4; Length 348;
Best Local Similarity 51.1%; Pred. No. 6.6e-68;
Matches 178; Conservative 38; Mismatches 106; Indels 26; Gaps 7;
QY 11 RVPLNSRESLSGETAQVYKTDNNVYQVDSKSFNDRVFGHNETTKNYEE 64
Db 1 RCPLNREINDGSCVQVPPWTGYKTENHGEHDSPHKSFTHVFWNCTQEDVYDT 60
QY 65 IAAPIIDSALQGVNGTIFAYGQTASGKTYTMMG---SEDLGVIPRAIHDIFOKIKKFP 120
Db 61 VAHPVDDCFHGYNCITFAFGTSGKTYTMMGPGGHPDHMGIIPCCHDIFDRIDKFQ 120
QY 121 DRE---FLRVSYMEIYNETITDLCCTQKPKPIIREDVARNVYVADLVEEVVYTSSE 176
Db 121 EKDHDFHVVHVSMEIYNEEYDILLCPNQHMKPLNIEHPNMGPPYVQGTETFEHVCYED 180
QY 177 ALKWKTKSRHGETKQNRSSRSHTIFRMILESREKGEPSNCEGSVKYSHLNLVDLA 236
Db 181 ACHWQGNKRRHVAATNMDHSSRSHTIFTHVEQRHK---QCDHVCCHSKNVLVDLA 236
QY 237 GSERAAQTGAAGVRLKEGCNINRSLFILQGVIKKLSGQV-----GGFINYRDSKLTIRI 290
Db 237 GSERVNRGTGAEGVRLKEGCNINRSLTILGNVINALADGQTKYMGHGHIPYRDSKLTWL 296
QY 291 LQNSLGNPKTRIICTIPV--SFEDELALQFASAKYMKNTPYNEV 336
Db 297 LQDSLGNCKTCMIACIWPADWNYEETLSTLRYADRAKNKKEPQINE 344

RESULT 6
ABU53125
ID ABU53125 standard; protein; 366 AA.
XX AC ABU53125;
XX DT 15-APR-2003 (first entry)
XX DE Intracellular trafficking-associated DKFZphtes3_26g22 homologue #2.
XX KW Human; gene therapy; vaccine; disease treatment; detection.
XX OS Homo sapiens.
XX PN WO200112659-A2.
XX PD 22-FEB-2001.
XX PF 18-AUG-2000; 2000WO-IB001496.
XX PR 18-AUG-1999; 99US-0149499P.
XX PR 28-SEP-1999; 99US-0156503P.
XX PA (GEHU-) GERMAN HUMAN GENOME PROJECT.
XX PI Wiemann S;
XX DR WPI; 2001-327840/34.
XX PT Nucleic acids having the sequences of clones isolated from libraries of different human tissues, useful in recombinant DNA methodologies.
XX PS Example III; Page 745; 1095pp; English.
XX PS This invention describes novel polynucleotides and polypeptides isolated from human cDNA libraries which can be used for gene therapy or in vaccines. The polynucleotides of the invention and antibodies encoded by them may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate polypeptide expression. The products of the invention may also be used to identify modulators of expression and activity and to down regulate expression and activity. The antibodies of the invention may also be used as diagnostic agents for detecting the presence of polypeptides in samples. This sequence represents a homologue of a polypeptide described in the disclosure of the invention

CC Presence of polypeptides in samples. This sequence represents a homologue
 CC of a polypeptide described in the disclosure of the invention
 XX
 SQ Sequence 366 AA;

Query Match 44.8%; Score 781; DB 4; Length 366;
 Best Local Similarity 51.1%; Pred. No. 7.1e-68;
 Matches 178; Conservative 38; Mismatches 106; Indels 26; Gaps 7;

QY 11 RVRPLNSREESLGETAQVYK-----KTDNNVIYQVDSKSFNDFRVFHNQETTKNYYE 64
 Db 1 RCRPLNERINDGSCCVQWPPWGTGKTVHNGHEGDSPHKSFTHVFWNCTQEDVYDT 60
 QY 65 IAAPILDSAIQYNGTIFAYGOTASGKTYTMMG-----SEDLGLVPIRAIHDIPOKIKKFP 120
 Db 61 VAHPVDDCFHGYNCITIFAYGOTGSKTYTMMGPGGEHPDHGIIIPRCCHDIFDRIDKQ 120
 QY 121 DRE---FLRVSYMEIYNETITDILLC-GTQMKPLIREDVNRNYYVADLVEEYVYSEM 176
 Db 121 EKDHDFHWKCSYMEIYNEEYDILCPNPQMKPLIHEHPNMPYVQCTEFHVCSEYD 180
 QY 177 ALKWTIKGSRHYGKTKNQSRSSHTIFRMILESEKGEPSNCEGSKVYSHLNLVDLA 236
 Db 181 ACHWVQGNKRVAAATNNDSSRSHTITFHVQRHK-----QCDEHYCHSKMNLVDLA 236
 QY 237 GSERAAQTGAAGVRLKEGNCINRSLFILGOVIKLSDGQV-----GGFINYRDSKLTRI 290
 Db 237 GSERNRTGAEGORLKEGNCINQSLTLGNVINADGQTKYMGHGHIPYRDSKLTRL 296
 QY 291 LQNSLGGNPKTRIICTITPV--SFDETLTALQFASHTAKYMNTPYVNE 336
 Db 297 LQDSLGGNCKTMIACIWPADWNBYETLTLRYADRAKNKPKQINE 344

RESULT 7
 ABB62322
 ID ABB62322 standard; protein; 2013 AA.
 AC ABB62322;
 DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 13758.
 DE Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 KW Drosophila melanogaster.
 OS
 XX WO200171042-A2.
 XX
 XX 27-SEP-2001.
 XX
 XX 23-MAR-2001; 2001WO-US009231.
 XX
 XX 23-MAR-2000; 2000US-0191637P.
 XX
 XX 11-JUL-2000; 2000US-00614150.
 XX
 XX (PEKE) PE CORP NY.
 XX
 XX Venter JC, Adams M, Li PWD, Myers EW;
 XX
 XX WPI; 2001-656860/75.
 XX
 XX N-PSDB; ABL06425.

PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.
 XX
 XX Disclosure; SEQ ID NO 13758; 21pp + Sequence Listing; English.
 XX
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 2013 AA;

Query Match 40.5%; Score 707; DB 4; Length 2013;
 Best Local Similarity 49.4%; Pred. No. 1.8e-59;
 Matches 166; Conservative 51; Mismatches 103; Indels 16; Gaps 8;

QY 5 AVAVCVRVPLNSREESLGETAQVYKWT-DNNVIYQVDSKSFNDFRVFHNQETTKNYY 62
 Db 8 SIQVCIKVAPCEPGLTSL-----WQVKERRSIHLADSHAEPYVDFYVDEGASNQEVF 60
 QY 63 BEIAAPIIDSAIQYNGTIFAYGOTASGKTYTMMGSEDLGLVPIRAIHDIPOKIKKFPDR 122
 Db 61 DMWAKHIVHACQMGNGTIFAYGOTSSGKTYTMMGDEQNPQGVMLAAKEIFQOISSETER 120
 QY 123 EELLRVSYMEIYNETITDILLCGTQMKPLIREDVNRNYYVADLVEEYVYSEM-ALKWI 181
 Db 121 DFLAVGVIEIYNEKIYDILL--NKNQDLKHESNGIVNVN--CEECITSEVDLLRL 176
 QY 182 TGKESRHYGKTKNQSRSSHTIFRMILESEKGEPSNCEGSKVYSHLNLVDLAGSERA 241
 Db 177 CLGNKERTVGETNMNERSRSRSHAIKFIIESR-KSDHSD-DDAVIQSVNLVDLAGSERA 234
 QY 242 AQTGAAGVRLKEGNCINRSLFILGOVIKLSDGQVGGFINYRDSKLTRLQNSLGGNPKT 301
 Db 235 DOTGARGARLKEGGHINKSLFLSNVIKLSNADNRFTNYRDSKLTRLQASLGGNAFT 294
 QY 302 RIICITPVSFDETLTALQFASHTAKYMNTPYVNEV 337
 Db 295 SIICITKESIMEESQSTLSFATRAKKIRIKPOVNEV 330

RESULT 8
 ABB70992
 ID ABB70992 standard; protein; 473 AA.
 AC ABB70992;
 XX
 XX 10-DEC-2002 (first entry)

XX Human target protein.
 XX
 XX Microtubule stimulated ATPase; cellular proliferation; cancer; AIDS;
 KW hyperplasia; restenosis; cardiac hypertrophy; immune disorder; HIV;
 KW inflammation; autoimmune disease; arthritis; graft rejection; psoriasis;
 KW inflammatory bowel disease; proliferation; medical procedure; surgery;
 KW human immunodeficiency virus; acquired immunodeficiency syndrome;
 KW angioplasty; human; Hskif; kinesin family.
 XX
 OS Homo sapiens.
 XX
 XX US6440684-B1.
 XX
 XX 27-AUG-2002.
 XX
 XX 12-JUN-2000; 2000US-00592054.
 XX
 XX 12-JUN-2000; 2000US-00592054.
 XX
 XX (CYTO-) CYTOKINETICS INC.
 XX
 XX Beraud C, Finer JT, Sakowicz R, Wood KW;
 XX WPI; 2002-711529/77.
 XX
 XX N-PSDB; ABB55162.

XX Screening for modulators of target protein having microtubule stimulated
PT ATPase activity e.g. kinesin family of protein, useful for treating
PT cancer, psoriasis, arthritis, human immunodeficiency virus (HIV)
XX infection.
XX
XX Claim 3; Fig 6; 34pp; English.
XX
XX The present invention relates to a new method of screening modulators of
CC target protein with microtubule stimulated ATPase activity. The method
CC involves contacting the target protein with an agent at 1st and 2nd
CC concentrations and determining the level of activity (e.g. binding or
CC ATPase activity) of target protein, where a difference between levels of
CC activity of target protein contacted with 1st and 2nd concentrations of
CC an agent indicates that an agent modulates activity of target protein.
CC The invention can be used for screening for modulators of target protein
CC having microtubule stimulated ATPase activity. The compounds identified
CC by method of the invention are useful for treating cellular proliferation
CC including cancer, hyperplasia, restenosis, cardiac hypertrophy, immune
CC disorders and inflammation. The compounds identified by the method are
CC also useful for treating autoimmune disease, arthritis, graft rejection,
CC inflammatory bowel disease, proliferation induced by medical procedures,
CC e.g. surgery, angioplasty etc. The compounds are also useful for treating
CC psoriasis. The compounds are useful for inhibiting human immunodeficiency
CC virus (HIV) and thus treating acquired immunodeficiency syndrome (AIDS).
CC The present amino acid sequence represents the human target protein of
CC the invention
XX
XX
XX Sequence 473 AA;
Query Match 38.3%; Score 668; DB 5; Length 473;
Best Local Similarity 44.7%; Pred. No. 1.4e-56;
Matches 152; Conservative 52; Mismatches 120; Indels 16; Gaps 5;
QY 6 VAVCVVRPLNSRBSLG-ETAQVYWKTDNNVYQVDSKSFNDRVPHGNETKNVYEE 64
DB 8 VRVALRCRPLVPKEISEGCMCLSFVPGEQVQVVGTD--KSFTYDFVDPSTEQEVFNT 65
QY 65 IAAPIIDSAIOGYNGTIFAYGQTASGKTYTMWG-----SEDLHGVIPRAIHDFQKIK 117
DB 66 AVAPLKGKFGYNATVLAQYGTGSGKTYSMGAYTAQENETPGVPIVQLLFKEID 125
QY 118 KPPDREFLLRVSYMEIYNETIDLLCGTQKMKPLIIRDNRNVTADLTREVVTSEMA 177
DB 126 KKSDFEFTLKVSYLEIYNEEILDLCPREKAQINIREDPKKGIVGLTEKTVLVALDT 185
QY 178 LKWTGKSRHYGTAKNQRSSHTIFRMLESREKGEPSNCSGVKVSHLNVLADLAG 237
DB 186 VSCLEQGNNSRTVASTAMNSQSSRSHAFITISLEQRKKS-----KNSFRSKHLHVLADLAG 241
QY 238 SERAQTGAAGVRLKEGNCINRSLFIQGVIKKLSGQVGGFYNRDSKLRILQNSLGG 297
DB 242 SERQKTKVAGDRLKEGININRGLICLGNVISALGDDKGGFVPRYDSKLRILQNSLGG 301
QY 298 NPKTRIICTITPV--SPDRTLALOFASAKYMKNTPVYN 335
DB 302 NSHTLMIAVCSPADSNLEETLNTLYADRAKIKKNKPIVN 341

RESULT 9
ID ABG70991
XX ABG70991 standard; protein; 522 AA.

AC ABG70991;

XX 10-DEC-2002 (first entry)

DE Human Hskif4 construct protein.

XX Microtubule stimulated ATPase; cellular proliferation; cancer; AIDS;
KW hyperplasia, restenosis; cardiac hypertrophy; immune disorder; HIV;
XX inflammation; autoimmune disease; arthritis; graft rejection; psoriasis;
KW inflammatory bowel disease; proliferation; medical procedure; surgery;

KW human immunodeficiency virus; acquired immunodeficiency syndrome;
KW angioplasty; human; Hskif4; kinesin family.
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FT Region 1..22 "N-terminal T7 epitope"
FT Region 496..516
FT Region /note= "C-terminal myc epitope"
FT Region 517..522
FT Region /note= "6-histidine residues at C-terminus"
XX
XX US6440684-B1.
XX 27-AUG-2002.
XX 12-JUN-2000; 2000US-00592054.
XX 12-JUN-2000; 2000US-00592054.
XX (CYTO-) CYTOKINETICS INC.
XX Beraud C, Finer JT, Sakowicz R, Wood KW;
XX WPT; 2002-711529/77.
XX N-PSDB; ABS55161.
XX
XX Screening for modulators of target protein having microtubule stimulated
PT ATPase activity e.g. kinesin family of protein, useful for treating
PT cancer, psoriasis, arthritis, human immunodeficiency virus (HIV)
PT infection.
XX
XX Claim 2; Fig 4; 34pp; English.
XX
XX The present invention relates to a new method of screening modulators of
CC target protein with microtubule stimulated ATPase activity. The method
CC involves contacting the target protein with an agent at 1st and 2nd
CC concentrations and determining the level of activity (e.g. binding or
CC ATPase activity) of target protein, where a difference between levels of
CC activity of target protein contacted with 1st and 2nd concentrations of
CC an agent indicates that an agent modulates activity of target protein.
CC The invention can be used for screening for modulators of target protein
CC having microtubule stimulated ATPase activity. The compounds identified
CC by method of the invention are useful for treating cellular proliferation
CC including cancer, hyperplasia, restenosis, cardiac hypertrophy, immune
CC disorders and inflammation. The compounds identified by the method are
CC also useful for treating autoimmune disease, arthritis, graft rejection,
CC inflammatory bowel disease, proliferation induced by medical procedures,
CC e.g. surgery, angioplasty etc. The compounds are also useful for treating
CC psoriasis. The compounds are useful for inhibiting human immunodeficiency
CC virus (HIV) and thus treating acquired immunodeficiency syndrome (AIDS).
CC The present amino acid sequence represents the human Hskif4 (kinesin
XX family) construct protein of the invention
XX
XX Sequence 522 AA;

Query Match 38.3%; Score 668; DB 5; Length 522;
Best Local Similarity 44.7%; Pred. No. 1.6e-56;
Matches 152; Conservative 52; Mismatches 120; Indels 16; Gaps 5;

QY 6 VAVCVVRPLNSRBSLG-ETAQVYWKTDNNVYQVDSKSFNDRVPHGNETKNVYEE 64
DB 30 VRVALRCRPLVPKEISEGCMCLSFVPGEQVQVVGTD--KSFTYDFVDPSTEQEVFNT 87
QY 65 IAAPIIDSAIOGYNGTIFAYGQTASGKTYTMWG-----SEDLHGVIPRAIHDFQKIK 117
DB 66 AVAPLKGKFGYNATVLAQYGTGSGKTYSMGAYTAQENETPGVPIVQLLFKEID 147
QY 118 KPPDREFLLRVSYMEIYNETIDLLCGTQKMKPLIIRDNRNVTADLTREVVTSEMA 177
DB 148 KKSDFEFTLKVSYLEIYNEEILDLCPREKAQINIREDPKKGIVGLTEKTVLVALDT 207

Qy	178	LKWITKGSKSRHYGETKKNORSRSHTIPRMILESKREKGEPCNCEGSVKVSHLNLVDLAG	237
Db	208	VSCLEQGNNSRTVASTAMNSQSSRSHAIFTISLEQRKKS-----KNSSFRSKLHLVDLAG	263
Qy	238	SERAAQTGAAGVRLKEGCNINRSLFTIGOVIKKLDGQGVGFNYRDSKLTILQNSLGG	297
Db	264	SERQKTKYAGDRLKEGINIRGLICLGNVISALGDDKKGGFVPYRDSKLTLLQDSLGG	323
Qy	298	NPKTRIITITPV--SFDETILALQFASAKYMKNTPYVN	335
Db	324	NSHTILMIACVSPADSNLBETLNTLRVADRARKIKNKPIVN	363
RESULT 10			
ADD49938			
ID	ADD49938 standard; protein; 1232 AA.		
XX	AC	ADD49938;	
XX	AC		
DT	15-JAN-2004 (first entry)		
XX			
DE	Human lung specific tumour antigen L 1447p.		
XX			
KW	Human; lung cancer antigen; cytostatic; lung cancer; gene therapy;		
KW	vaccine; T-cell; tumour.		
XX			
OS	Homo sapiens.		
XX			
PN	US2003194764-A1.		
XX			
PD	16-OCT-2003.		
XX			
PF	04-APR-2002; 2002US-00116712.		
XX			
PR	05-APR-2001; 2001US-0282289P.		
PR	05-OCT-2001; 2001US-0327511P.		
XX			
PA	(CORI-) CORIXA CORP.		
XX			
PI	Bangor CS, Switzer A;		
XX			
DR	WPI; 2003-844452/78.		
DR	N-PSDB; ADD49936.		
XX			
PT	New isolated polypeptides and polynucleotides useful for diagnosing,		
PT	preventing and treating cancer, particularly lung cancer.		
XX			
PS	Claim 1; SEQ ID NO 670; 250pp; English.		

patient), a method for stimulating and/or expanding T-cells specific for a tumour protein (comprising contacting T-cells with the above polypeptide, polynucleotide or antigen-presenting cells that express the polynucleotide, under conditions and for a time sufficient to permit the stimulation and/or expansion of T-cells), an isolated T-cell population comprising T-cells prepared by the method, a composition comprising a first component selected from physiological carriers and immunostimulants, and a second component selected from the above polypeptide, polynucleotide, antibody, fusion protein, T-cell population and antigen-presenting cells that express the above polypeptide, stimulating an immune response in a patient (comprising administering to the patient the above composition) treating lung cancer in a patient (comprising administering to the patient the above composition and a diagnostic kit (comprising: at least one oligonucleotide cited above; or at least one antibody cited above and a detection reagent, where the detection reagent comprises a reporter group). The composition and methods are useful in diagnosing, preventing and treating cancer, particularly lung cancer. The present sequence is a lung cancer-associated antigen of the invention.

QY	178	LKWITKGEKSRHYGVTKNORSSRSHHTIFRMILESRKEGPEPNCBGSKVVKSHLNUVDLAG	233
Db	208	VSCLEQGNNSRTVASTAMNSQSRHAFTTISLEQRKKS-----KNSSPRSKLHLVDLAG	263
QY	238	SERAAQTGAAGVRLKEGCNINRSLFILGOVIKKLSDGOVGGFINVRDSKLTILQNSLGG	297
Db	264	SERQKTKAEGDRLKEGININRGLLCLGNVISALGDDKKGFFVYRDSKLTLLQDSLGG	323
QY	298	NPKTRIILCTIPV--SFDETLTALQFASTAKYMKNTIPPYN	335
Db	324	NSHTLMIAVCSPADSNLBETLTNLTLYADRARKIKNKPIN	363
RESULT 10			
ADD49938			
ID	ADD49938	standard; protein; 1232 AA.	
AC	ADD49938;		
XX	XX	15-JAN-2004 (first entry)	
DT	DT	Human lung specific tumour antigen L 1447p.	
DE	DE		
KW	KW	Human; lung cancer antigen; cytostatic; lung cancer; gene therapy;	
KW	KW	vaccine; T-cell; tumour.	
XX	XX		
OS	OS	Homo sapiens.	
PN	PN	US2003194764-A1.	
XX	XX		
PD	PD	16-OCT-2003.	
PF	PF		
XX	XX	04-APR-2002; 2002US-00116712.	
PR	PR		
XX	XX	05-APR-2001; 2001US-0282289P.	
PR	PR		
XX	XX	05-OCT-2001; 2001US-0327511P.	
XX	XX		
PA	PA	(CORI-) CORIXA CORP.	
XX	XX		
PI	PI	Bangor CS, Switzer A;	
XX	XX		
DR	DR	WPI; 2003-844452/78.	
XX	XX	N-PSDB; ADD49936.	
DR	DR		
XX	XX		
PT	PT	New isolated polypeptides and polynucleotides useful for diagnosing,	
PT	PT	preventing and treating cancer, particularly lung cancer.	
XX	XX		
PS	PS	Claim 1; SEQ ID NO 670; 250pp; English.	
XX	XX		
CC	CC	The invention relates to an isolated polynucleotide (a) comprising any of	
CC	CC	the 666 fully defined nucleotide sequences appearing as ADD49269 -	
CC	CC	ADD49935, ADD49936 and ADD49938, complements of (a); sequences of at	
CC	CC	least 20 contiguous residues of (a); sequences that hybridise to (a)	
CC	CC	under highly stringent conditions; sequences having at least 75 or 90%	
CC	CC	identity to (a); or degenerate variants of (a). Also included are an	
CC	CC	isolated polypeptide (b) (comprising: sequences encoded by the new	
CC	CC	polynucleotide; any of the 4 amino acid sequences fully defined in the	
CC	CC	specification; or sequences having at least 70 or 90% identity to the	
CC	CC	sequence in (a) or (b)), an expression vector comprising the above	
CC	CC	polynucleotide operably linked to an expression control sequence, a host	
CC	CC	cell transformed or transfected with the above expression vector, an	
CC	CC	isolated antibody, or its antigen-binding fragment, that specifically	
CC	CC	binds to the above polypeptide, an oligonucleotide that hybridises to the	
CC	CC	above-mentioned nucleotide sequences under highly stringent conditions, a	
CC	CC	fusion protein comprising at least one polypeptide cited above, the	
CC	CC	presence of a cancer in a patient (comprising: obtaining a biological	
CC	CC	sample from the patient; contacting the biological sample with a binding	
CC	CC	agent that binds to the polypeptide, or with the oligonucleotide cited	
CC	CC	above; detecting in the sample an amount of the polypeptide that binds to	
CC	CC	the binding agent, or an amount of a polynucleotide that hybridises to	
CC	CC	the oligonucleotide; and comparing the amount of polypeptide, or	
CC	CC	polynucleotide that hybridises to the oligonucleotide, to a predetermined	
CC	CC	cut-off value and then determining the presence of a cancer in the	

XX PD 06-MAR-2003.

XX PF 23-AUG-2002; 2002WO-GB0003892.

XX PR 23-AUG-2001; 2001GB-00020558.

XX PR 05-OCT-2001; 2001GB-00024037.

XX PA (OXFO-) OXFORD BIOMEDICA UK LTD.

XX PI Kingsman SM, White J, Ward NR, Harris RA, Naylor S, Mundy CR;

XX PD WPI; 2003-290046/28.

XX DR N-PSDB; ADD18925.

XX

PT New substantially purified polypeptide, useful for diagnosing or treating

PT a hypoxia-regulated condition, such as cancer, ischemia, reperfusion

PT injury, retinopathy, pre-eclampsia, atherosclerosis, inflammation, or

PT wound healing.

XX

PS Claim 25; SEQ ID NO 413; 424pp; English.

XX

CC This invention relates to novel human genes and gene product which are

CC implicated in certain disease states. Compounds which modulate the

CC proteins of the invention may have cytostatic, antiinflammatory,

CC ophthalmological, antiarteriosclerotic or vulnerary activities. The

CC sequences of the invention may be useful for gene therapy. The invention

CC may be useful for diagnosing or treating a hypoxia-regulated condition,

CC such as tumorigenesis, angiogenesis, apoptosis, inflammation,

CC erythropoiesis, or the biological response to hypoxia conditions

CC including processes such as glycolysis, gluconeogenesis, glucose

CC transportation, catecholamine synthesis, iron transport or nitric oxide

CC synthesis. The disease includes cancer, ischaemic conditions, reperfusion

CC injury, retinopathy, neonatal stress, pre-eclampsia, atherosclerosis,

CC inflammatory conditions or wound healing. The present sequence is that of

CC a disease related protein of the invention.

XX

XX Sequence 1232 AA;

Query Match 38.3%; Score 668; DB 7; Length 1232;

Best Local Similarity 44.7%; Pred. No. 5.9e-56;

Matches 152; Conservative 52; Mismatches 120; Indels 16; Gaps 5;

QY 6 VAVCVRPRLNRSRESLG-ETAQVYVTKDNNVIYQVQDGSKFDFRPHGNKTNVYEE 64

Db 10 VRVALRCEPLVPKEISECQCLSFVPGEPQVVGTD--KSFTYDFVDPSTEQEVEVT 67

QY 65 IAAPLIDSAIOGYNGTIFAYQGTASGKTYTWMG-----SEDLHGVIPRAIHIFQKIK 117

Db 68 AVAPLIKGVFGYNATVLAYGQTSGKTYSMGGAYTAQENEPTVGVIPRVLIQLFKRID 127

QY 118 KFPDFEFLRVSYMEIYNETITDLCTQKMKPLIREDVNRNVYVADLTREVVYVYSEMA 177

Db 128 KKSDFEFLRVSYMEIYNETITDLCTQKMKPLIREDVNRNVYVADLTREVVYVYSEMA 177

QY 178 LKWTIKGKSHYGTCKMNRSSRSHITFRMILSRKGEPSNCEGSKVSHNLVLDIAG 237

Db 188 VSCLEQGNNSRTVASTAMNSQSRSHAFTISLEQRKSD---KNSSFRSKLHLVDLAG 243

QY 238 SERAAQTGAAGVRLKEGNNKNSLFTLGQVINKLSGQGVGFINTVRDLSKLTILQNSLGG 297

Db 244 SERQKTKAEGDRLKEGININRGLLCLGNVISALGDDKGGFVPRYDRSKLTRELLQDSLGG 303

QY 298 NPKTILICTIPV--SFDETLTALQFASAKYMKNTPVYN 335

Db 304 NSHTLMIACVSPADSNLEETLNLRYADARKIKNKPVYN 343

RESULT 12

ADD49937

ID ADD49937 standard; protein; 1232 AA.

XX

AC ADD49937;

XX DT 15-JAN-2004 (first entry)

XX DE Human kinesin family member 4A.

XX

XX KW Human; lung cancer antigen; cytostatic; lung cancer; gene therapy;

XX KW vaccine; T-cell; tumour.

XX OS Homo sapiens.

XX PN US2003194764-A1.

XX PD 16-OCT-2003.

XX PF 04-APR-2002; 2002US-00116712.

XX PR 05-APR-2001; 2001US-0282289P.

XX PR 03-OCT-2001; 2001US-0327511P.

XX PA (CORI-) CORIXA CORP.

XX PI Bangur CS, Switzer A.

XX

XX WPI; 2003-844452/78.

XX DR N-PSDB; ADD49935.

XX

PT New isolated polypeptides and polynucleotides useful for diagnosing,

PT preventing and treating cancer, particularly lung cancer.

XX

XX Example 3; SEQ ID NO 669; 250pp; English.

XX

CC The invention relates to an isolated polynucleotide (a) comprising any of

CC the 666 fully defined nucleotide sequences appearing as ADD49289

CC ADD49935, ADD49936 and ADD49938, complements of (a); sequences of at

CC least 20 contiguous residues of (a); sequences that hybridise to (a)

CC under highly stringent conditions; sequences having at least 75 or 90%

CC identity to (a); or degenerate variants of (a). Also included are an

CC isolated polypeptide (b) (comprising: sequences encoded by the new

CC polynucleotide; any of the 4 amino acid sequences fully defined in the

CC specification; or sequences having at least 70 or 90% identity to the

CC sequence in (a) or (b)), an expression vector comprising the above

CC polynucleotide operably linked to an expression control sequence, a host

CC cell transformed or transfected with the above expression vector, an

CC isolated antibody, or its antigen-binding fragment, that specifically

CC binds to the above polypeptide, an oligonucleotide that hybridises to the

CC above-mentioned nucleotide sequences under highly stringent conditions, a

CC fusion protein comprising at least one polypeptide cited above, detecting

CC the presence of a cancer in a patient (comprising: obtaining a biological

CC sample from the patient; contacting the biological sample with a binding

CC agent that binds to the polypeptide, or with the oligonucleotide cited

CC above; detecting in the sample an amount of the polypeptide that binds to

CC the binding agent, or an amount of a polynucleotide that hybridises to

CC the oligonucleotide; and comparing the amount of polypeptide, or

CC polynucleotide that hybridises to the oligonucleotide, to a predetermined

CC cut-off value and then determining the presence of a cancer in the

CC patient), a method for stimulating and/or expanding T-cells specific for

CC a tumour protein (comprising: contacting T-cells with the above

CC polypeptide, polynucleotide or antigen-presenting cells that express the

CC polynucleotide, under conditions and for a time sufficient to permit the

CC stimulation and/or expansion of T-cells), an isolated T-cell population

CC comprising T-cells prepared by the method, a composition comprising a

CC first component selected from physiological carriers and

CC immunostimulants, and a second component selected from the above

CC polypeptide, polynucleotide, antibody, fusion protein, T-cell population

CC and antigen-presenting cells that express the above polypeptide,

CC stimulating an immune response in a patient (comprising administering to

CC the patient the above composition) treating lung cancer in a patient

CC (comprising administering to the patient the above composition and a

CC diagnostic kit (comprising: at least one oligonucleotide cited above; or

CC at least one antibody cited above and a detection reagent, where the

CC detection reagent comprises a reporter group). The composition and

CC methods are useful in diagnosing, preventing and treating cancer,

CC particularly lung cancer. The present sequence is a lung cancer-

CC associated antigen of the invention.

XX Sequence 1232 AA;

Query Match 37.7%; Score 658; DB 7; Length 1232;
Best Local Similarity 44.7%; Pred. No. 5.6e-55;
Matches 152; Conservative 52; Mismatches 120; Indels 16; Gaps 6;

QY 6 VAVCVVRPLNSREESLG-ETAQVYWKTDNNVIYQVDGSKSFNDFRVFHNQNTKNVYEE 64
DB 10 VRVALRCRPLVPKEISEGCMCLSFVPGEQVVVGTD--KSTYDFVDFPSTQEVEFMT 67
QY 65 IAPIIDSAIQYNGTIFAYGQTASGKTYTMMG-----SEHLGVIPRAIHDFQKIK 117
DB 68 AVAPLKGKGVKYNATVLAQGTSKTYSMGAVTAEGNEPTGVIPRVQLLFKEID 127
QY 118 KFPDREFLLRVSYMEIYNETITDLCCTOKMKPLIREDVNRNRYVADLTVEVVYVTSMA 177
DB 128 KKSDFEFTLKVSLEYNEIIDLCPSEKQAQINIREDPKEGIKIVGITEKTVLVALDT 187
QY 178 LKWITKGEKSRHYGETKMNQSRSHITFRMILESEKGEPCNCEGSKVSHNLNVDLAG 237
DB 188 VSCLEQGNNSRTVA STMNSQSSRSHAFTI---SLEQKKSKDNSSFR-SKLHLVDLAG 243
QY 238 SERAAQTGAAGVRLKEGNCNINRSFLIGQVVKLSGQGVGFYNYRDSKLTIRLQNSLGG 297
DB 244 SERQKTKAEGDRLEKGININRGLLCLGNVISALGDDKKGFPYRDSKLTIRLQDSLGG 303
QY 298 NPKTRIICTITPV--SFDETLTALQFASTAKYMKNTPYVN 335
DB 304 NSHTLMIACVSPADSNLEETLTLRYADRARKINKPIVN 343

RESULT 13
ADD49932

ID ADD49932 standard; protein; 1232 AA.

AC ADD49932;

DT 15-JAN-2004 (first entry)

XX Human lung specific tumour antigen L 1477p.

XX Human; lung cancer antigen; cytostatic; lung cancer; gene therapy;
XX vaccine; T-cell; tumour.

OS Homo sapiens.

PN US2003194764-A1.

PD 16-OCT-2003.

XX 04-APR-2002; 2002US-00116712.

XX 05-APR-2001; 2001US-0282289P.

XX 05-OCT-2001; 2001US-0327511P.

XX (CORI-) CORIXA CORP.

XX Bangor CS, Switzer A;

XX WPI; 2003-844452/78.

XX N-PSDS; ADD49929.

XX New isolated polypeptides and polynucleotides useful for diagnosing,
XX preventing and treating cancer, particularly lung cancer.

XX Example 3; SEQ ID NO 664; 250pp; English.

XX The invention relates to an isolated polynucleotide (a) comprising any of
XX the 666 fully defined nucleotide sequences appearing as ADD49269 -
XX ADD49935, ADD49936 and ADD49938, complements of (a); sequences of (a)
XX least 20 contiguous residues of (a); sequences that hybridise to (a)

CC under highly stringent conditions; sequences having at least 75 or 90%
CC identity to (a); or degenerate variants of (a). Also included are an
CC isolated polypeptide (b) comprising: sequences encoded by the new
CC polynucleotide; any of the 4 amino acid sequences fully defined in the
CC specification; or sequences having at least 70 or 90% identity to the
CC sequence in (a) or (b), an expression vector comprising the above
CC polynucleotide operably linked to an expression control sequence, a host
CC cell transformed or transfected with the above expression vector, a host
CC isolated antibody, or its antigen-binding fragment, that specifically
CC binds to the above polypeptide, an oligonucleotide that hybridises to the
CC above-mentioned nucleotide sequences under highly stringent conditions, a
CC fusion protein comprising at least one polypeptide cited above, detecting
CC the presence of a cancer in a patient (comprising: obtaining a biological
CC sample from the patient; contacting the biological sample with a binding
CC agent that binds to the polypeptide, or with the oligonucleotide cited
CC above; detecting in the sample an amount of the polypeptide that binds to
CC the binding agent, or an amount of a polynucleotide that hybridises to
CC the oligonucleotide; and comparing the amount of polypeptide, or
CC polynucleotide that hybridises to the oligonucleotide, to a predetermined
CC cut-off value and then determining the presence of a cancer in the
CC patient), a method for stimulating and/or expanding T-cells specific for
CC a tumour protein (comprising contacting T-cells with the above
CC polypeptide, polynucleotide or antigen-presenting cells that express the
CC polynucleotide, under conditions and for a time sufficient to permit the
CC stimulation and/or expansion of T-cells), an isolated T-cell population
CC comprising T-cells prepared by the method, a composition comprising a
CC first component selected from physiological carriers and
CC immunostimulants, and a second component selected from the above
CC polypeptide, polynucleotide, antibody, fusion protein, T-cell population
CC and antigen-presenting cells that express the above polypeptide,
CC stimulating an immune response in a patient (comprising administering to
CC the patient the above composition) treating lung cancer in a patient
CC (comprising administering to the patient the above composition and a
CC diagnostic kit (comprising: at least one oligonucleotide cited above; or
CC at least one antibody cited above and a detection reagent, where the
CC detection reagent comprises a reporter group). The composition and
CC methods are useful in diagnosing, preventing and treating cancer,
CC particularly lung cancer. The present sequence is a lung cancer-
XX associated antigen of the invention.

XX Sequence 1232 AA;

Query Match 37.7%; Score 658; DB 7; Length 1232;
Best Local Similarity 44.7%; Pred. No. 5.6e-55;
Matches 152; Conservative 52; Mismatches 120; Indels 16; Gaps 6;

QY 6 VAVCVVRPLNSREESLG-ETAQVYWKTDNNVIYQVDGSKSFNDFRVFHNQNTKNVYEE 64
DB 10 VRVALRCRPLVPKEISEGCMCLSFVPGEQVVVGTD--KSTYDFVDFPSTQEVEFMT 67
QY 65 IAPIIDSAIQYNGTIFAYGQTASGKTYTMMG-----SEHLGVIPRAIHDFQKIK 117
DB 68 AVAPLKGKGVKYNATVLAQGTSKTYSMGAVTAEGNEPTGVIPRVQLLFKEID 127
QY 118 KFPDREFLLRVSYMEIYNETITDLCCTOKMKPLIREDVNRNRYVADLTVEVVYVTSMA 177
DB 128 KKSDFEFTLKVSLEYNEIIDLCPSEKQAQINIREDPKEGIKIVGITEKTVLVALDT 187
QY 178 LKWITKGEKSRHYGETKMNQSRSHITFRMILESEKGEPCNCEGSKVSHNLNVDLAG 237
DB 188 VSCLEQGNNSRTVA STMNSQSSRSHAFTI---SLEQKKSKDNSSFR-SKLHLVDLAG 243
QY 238 SERAAQTGAAGVRLKEGNCNINRSFLIGQVVKLSGQGVGFYNYRDSKLTIRLQNSLGG 297
DB 244 SERQKTKAEGDRLEKGININRGLLCLGNVISALGDDKKGFPYRDSKLTIRLQDSLGG 303
QY 298 NPKTRIICTITPV--SFDETLTALQFASTAKYMKNTPYVN 335
DB 304 NSHTLMIACVSPADSNLEETLTLRYADRARKINKPIVN 343

RESULT 14
ABG70990

ABG70990 standard; protein; 1232 AA.
ABG70990;
10-DEC-2002 (first entry)
Human HsKif4 protein.
Microtubule stimulated ATPase; cellular proliferation; cancer; AIDS;
hyperplasia; restenosis; cardiac hypertrophy; immune disorder; HIV;
inflammation; autoimmune disease; arthritis; graft rejection; psoriasis;
inflammatory bowel disease; proliferation; medical procedure; surgery;
human immunodeficiency virus; acquired immunodeficiency syndrome;
angioplasty; human; HsKif4; kinesin family.
Homo sapiens.
Key Location/Qualifiers
Misc-difference 161..162 /note= "Encoded by TTAGAG"
Misc-difference 164..180 /note= "Encoded by TACAATGAAGAAATTGGATCTTCTAATGTC-
CATCTCGTGAGAAAGCTCAA"
US6440684-B1.
27-AUG-2002.
12-JUN-2000; 2000US-00592054.
12-JUN-2000; 2000US-00592054.
(CYTO-) CYTOKINETICS INC.
Beraud C, Finer JT, Sakowicz R, Wood KW;
WPI; 2002-711529/77.
N-PSDB; ABS55160.
Screening for modulators of target protein having microtubule stimulated
ATPase activity e.g. kinesin family of protein, useful for treating
cancer, psoriasis, arthritis, human immunodeficiency virus (HIV)
infection.
Claim 1; Fig 2; 34pp; English.
The present invention relates to a new method of screening modulators of
target protein with microtubule stimulated ATPase activity. The method
involves contacting the target protein with an agent at 1st and 2nd
concentrations and determining the level of activity (e.g. binding or
ATPase activity) of target protein, where a difference between levels of
activity of target protein contacted with 1st and 2nd concentrations of
an agent indicates that an agent modulates activity of target protein.
The invention can be used for screening for modulators of target protein
having microtubule stimulated ATPase activity. The compounds identified
by method of the invention are useful for treating cellular proliferation
including cancer, hyperplasias, restenosis, cardiac hypertrophy, immune
disorders and inflammation. The compounds identified by the method are
also useful for treating autoimmune disease, arthritis, graft rejection,
inflammatory bowel disease, proliferation induced by medical procedures,
e.g. surgery, angioplasty etc. The compounds are also useful for treating
psoriasis. The compounds are useful for inhibiting human immunodeficiency
virus (HIV) and thus treating acquired immunodeficiency syndrome (AIDS).
The present amino acid sequence represents the human HsKif4 (kinesin
family) protein of the invention
Sequence 1232 AA;
Query Match 37.4%; Score 652; DB 5; Length 1232;
Best Local Similarity 44.4%; Pred. No. 2.2e-54;
Matches 151; Conservative 52; Mismatches 121; Indels 16; Gaps 6;
6 VAVCVRVFLNGBESLG-ETAQVYWKTDNNVYQVDSKSPFDRVPHGNETTKNVEE 64

Db 10 VVALRCRLVPKELSEGQCMCLSFVGPQVVGTD--KSTYDFVDPSTEQBEVENT 67
QY 65 IAAPIIDSAIOGYNGTIFAYGQTASGKTYTMG-----SEDLGLVTPRAHDIFQKIK 117
Db 68 AVAPLIKGVKGYNAIVLAYGQTSGKTYSMGAYTAQENEPVGVIPRVQILLFKEID 127
QY 118 KPDPREFLRVSYMEIYNETITDLGCTQMKPLIITREDVNNVYVADLTTEVVYVSEMA 177
Db 128 KKSDFEFTLKVSLEYETINEEIIIDLLCPREKAQINIREDPKEGKIVGLTEKTVLVALDT 187
QY 178 LKMWITGKESRHYGETKMNORSRSHITIFRMILESEKGEPSNCEGSVKVSHLNVLDLAG 237
Db 188 VSCLEGQNNRPTVASTAMNSQSSRHAILTI---SLEQKKSKDKNSSFR-SKLHLVLDLAG 243
QY 238 SERAAQTGAAGVRLKEGCNINRSLFVLGQVIKLSGQVGGFVNYNDSKLTILQNSLGG 297
Db 244 SERQKTKAEGDRLKEGINIRGLLCLGNVISALGDDKGGFAPYRDSKLTLLQDSLGG 303
QY 298 NPKTRILICTIPV--SFDETLTALQFASTAKYMNTPYVN 335
Db 304 NSHTLMACVSPADSNLEETLNTLRVADRARKINRPIVN 343
RESULT 15
ABG70993
ID ABG70993 standard; protein; 1234 AA.
XX
AC ABG70993;
XX
DT 10-DEC-2002 (first entry)
XX
DE Human HsKif4b protein.
XX
KW Microtubule stimulated ATPase; cellular proliferation; cancer; AIDS;
hyperplasia; restenosis; cardiac hypertrophy; immune disorder; HIV;
inflammation; autoimmune disease; arthritis; graft rejection; psoriasis;
inflammatory bowel disease; proliferation; medical procedure; surgery;
human immunodeficiency virus; acquired immunodeficiency syndrome;
angioplasty; human; HsKif4b; kinesin family.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 850 /note= "Encoded by CGBCA"
FT FT
XX US6440684-B1.
XX
PD 27-AUG-2002.
XX
PF 12-JUN-2000; 2000US-00592054.
XX
PR 12-JUN-2000; 2000US-00592054.
XX
PA (CYTO-) CYTOKINETICS INC.
XX
PI Beraud C, Finer JT, Sakowicz R, Wood KW;
XX
DR WPI; 2002-711529/77.
XX
DR N-PSDB; ABS55163.
XX
PT Screening for modulators of target protein having microtubule stimulated
ATPase activity e.g. kinesin family of protein, useful for treating
cancer, psoriasis, arthritis, human immunodeficiency virus (HIV)
infection.
XX
PS Claim 1; Fig 2; 34pp; English.
XX
CC The present invention relates to a new method of screening modulators of
target protein with microtubule stimulated ATPase activity. The method
involves contacting the target protein with an agent at 1st and 2nd
concentrations and determining the level of activity (e.g. binding or
ATPase activity) of target protein, where a difference between levels of
activity of target protein contacted with 1st and 2nd concentrations of
an agent indicates that an agent modulates activity of target protein.
The invention can be used for screening for modulators of target protein
having microtubule stimulated ATPase activity. The compounds identified
by method of the invention are useful for treating cellular proliferation
including cancer, hyperplasias, restenosis, cardiac hypertrophy, immune
disorders and inflammation. The compounds identified by the method are
also useful for treating autoimmune disease, arthritis, graft rejection,
inflammatory bowel disease, proliferation induced by medical procedures,
e.g. surgery, angioplasty etc. The compounds are also useful for treating
psoriasis. The compounds are useful for inhibiting human immunodeficiency
virus (HIV) and thus treating acquired immunodeficiency syndrome (AIDS).
The present amino acid sequence represents the human HsKif4 (kinesin
family) protein of the invention
Sequence 1232 AA;
Query Match 37.4%; Score 652; DB 5; Length 1232;
Best Local Similarity 44.4%; Pred. No. 2.2e-54;
Matches 151; Conservative 52; Mismatches 121; Indels 16; Gaps 6;
6 VAVCVRVFLNGBESLG-ETAQVYWKTDNNVYQVDSKSPFDRVPHGNETTKNVEE 64

Search completed: July 29, 2004, 09:35:12
Job time : 21.4051 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 09:26:01 ; Search time 160.292 Seconds
(without alignments)
4694.096 Million cell updates/sec

Title: US-10-045-631B-88

Perfect score: 13329

Sequence: 1 MAEEGAVAVCVRRVPLNSRE.....SQPGMHASSGKDVPCKTQ 2663

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq_29Jan04.*

1: Geneseqp1980s.*

2: Geneseqp1990s.*

3: Geneseqp2000s.*

4: Geneseqp2001s.*

5: Geneseqp2002s.*

6: Geneseqp2003as.*

7: Geneseqp2003bs.*

8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	13305	99.8	2663	4 AAM39097	Aam39097 Human pol
2	13170	98.8	2688	4 AAM40883	Aam40883 Human pol
3	13131	98.5	2633	4 ABG06505	Abg06505 Novel hum
4	13115	27.1	2954	2 AAY01632	Aay01632 Amino aci
5	1545.5	11.6	2013	4 ABB62322	Abb62322 Drosophil
6	1286	9.6	1931	4 ABB61012	Abb61012 Drosophil
7	1183	8.9	240	5 AAU83007	Aau83007 Human hom
8	1127.5	8.5	3259	7 ADE56037	Ades6037 Human pro
9	1127.5	8.5	3259	7 ADE56033	Ades6033 Human pro
10	1091	8.2	2228	7 ABR61599	Abr61599 Human gol
11	1091	8.2	2230	6 ABU07445	Abu07445 Protein d
12	1091	8.2	2230	7 ABR61600	Abr61600 Human gol
13	1089	8.2	2252	7 ABR61602	Abr61602 Human gol
14	1089	8.2	3187	7 ADE56031	Ades6031 Rat Prote
15	1089	8.2	3187	7 ADE56035	Ades6035 Rat Prote
16	1089	8.2	3187	7 ADE56035	Ades6035 Rat Prote
17	1077.5	8.1	1388	5 AAU14400	Aau14400 Human kin
18	1077.5	8.1	1388	5 AAU79590	Aau79590 Human kin
19	1077.5	8.1	1388	6 ABR48222	Abr48222 Human bla
20	1077.5	8.1	1388	7 ADC35116	Adc35116 Human bre
21	1072.5	8.0	1388	7 ADB80468	Adb80468 Ovarian c
22	1064	8.0	3899	6 ABR92048	Abr92048 Human cer
23	1064	8.0	3917	6 ABR92050	Abr92050 Human cer
24	1058	7.9	3907	6 ABR92047	Abr92047 Human cer
25	1058	7.9	3925	6 ABR92049	Abr92049 Human cer

ALIGNMENTS

RESULT 1

AAM39097
ID AAM39097 standard; protein; 2663 AA.
XX AC AAM39097;
XX XX

22-OCT-2001 (first entry)
XX XX

Human polypeptide SEQ ID NO 2242.
XX XX

Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer;
peripheral nervous system; neuropathy; central nervous system; CNS;
Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
leukaemia.
XX XX

Homo sapiens.
XX XX

WO200153312-A1.
XX XX

26-JUL-2001.
XX XX

26-DEC-2000; 2000WO-US034263.
XX XX

23-DEC-1999; 99US-00471275.
XX XX

21-JAN-2000; 2000US-00488725.
XX XX

25-APR-2000; 2000US-0052317.
XX XX

20-JUN-2000; 2000US-00598042.
XX XX

19-JUL-2000; 2000US-00620312.
XX XX

03-AUG-2000; 2000US-00653450.
XX XX

14-SEP-2000; 2000US-00662191.
XX XX

19-OCT-2000; 2000US-00693036.
XX XX

29-NOV-2000; 2000US-00727344.
XX XX

(HYSE-) HYSEQ INC.
XX XX

Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX XX

Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
XX XX

Zhou P, Goodrich R, Drmanac RT;
XX XX

WPI: 2001-442253/47.
XX XX

N-PSDB; AAI58253.
XX XX

Novel nucleic acids and polypeptides, useful for treating disorders such
XX XX

as central nervous system injuries.
XX XX

Example 4; SEQ ID NO 2242; 10078pp; English.
XX XX

XX The invention relates to human nucleic acids (AA157798-AA161369) and the
CC encoded polypeptides (AAM3642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders. Note: The sequence data for this patent did not form
CC part of the printed specification
SQ
XX Sequence 2663 AA;
Query Match 99.8%; Score 13305; DB 4; Length 2663;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2659; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 1 MAEGAVAVCVVRPLNGREESLGETAQYVWKTNNVIYQVDSKSFNDRVFHGNETK 60
DB 1 MAEGAVAVCVVRPLNGREESLGETAQYVWKTNNVIYQVDSKSFNDRVFHGNETK 60
QY 61 NYVEETAAPIIDSAIQGYNGTIFAYGQTASGKTYTMMGSEDLGVIPRAIHDFQKIKKF 120
DB 61 NYVEETAAPIIDSAIQGYNGTIFAYGQTASGKTYTMMGSEDLGVIPRAIHDFQKIKKF 120
QY 121 POREFLLRVSYMEIYNETITDILCGTQKMKPLIIRREDVNRNVYVADLTSEVVYVTEMAUK 180
DB 121 POREFLLRVSYMEIYNETITDILCGTQKMKPLIIRREDVNRNVYVADLTSEVVYVTEMAUK 180
QY 181 WITKGEKSHYGETKMQNRSSSHITFRMILSREKGEPSNCEGSKVSHNLNLDLAGSE 240
DB 181 WITKGEKSHYGETKMQNRSSSHITFRMILSREKGEPSNCEGSKVSHNLNLDLAGSE 240
QY 241 RAAQTGAAGVRLKEGNCINRSIFILGOVTKLSDGQGVGFIVNRDSKLTFRILQNSLGGNP 300
DB 241 RAAQTGAAGVRLKEGNCINRSIFILGOVTKLSDGQGVGFIVNRDSKLTFRILQNSLGGNP 300
QY 301 KTRIICTITPVSFDETLTALQFASTAKYMKNTPYVNEVSTDEALLKRYRKEIMDLKKOLE 360
DB 301 KTRIICTITPVSFDETLTALQFASTAKYMKNTPYVNEVSTDEALLKRYRKEIMDLKKOLE 360
QY 361 EVSLETRAQAMEKDQALLEEKDLLOKVQNEKIENLTMIVTSSSLTQOELLKAKRKR 420
DB 361 EVSLETRAQAMEKDQALLEEKDLLOKVQNEKIENLTMIVTSSSLTQOELLKAKRKR 420
QY 421 VTWCLGKINKMNSYADQFNIPNTITTKTKLSINLLREIDRSVCSESDFVNTLDTLS 480
DB 421 VTWCLGKINKMNSYADQFNIPNTITTKTKLSINLLREIDRSVCSESDFVNTLDTLS 480
QY 481 EIEWNPATKLLNQENIESELNSLRADYDNLVLDYEQLRTEKEEMELKKEKNLDLFEAL 540
DB 481 EIEWNPATKLLNQENIESELNSLRADYDNLVLDYEQLRTEKEEMELKKEKNLDLFEAL 540
QY 541 ERKTKKQDMQLIHEISNLKLVKRVHVNQDLLENELSSKVELLRKEKDOIKKLOEYIDS 600
DB 541 ERKTKKQDMQLIHEISNLKLVKRVHVNQDLLENELSSKVELLRKEKDOIKKLOEYIDS 600
QY 601 QKLENTKMDLSYLSIESIEDPKQKOTLFDATVALDAKRESAFILSENLELKEKMKELAT 660
DB 601 QKLENTKMDLSYLSIESIEDPKQKOTLFDATVALDAKRESAFILSENLELKEKMKELAT 660
QY 661 TYQKEMENDIQLYQSLEAKKMQVDLEKELQSAFNEITKLTSLIDGKVPKDLNLELEG 720
DB 661 TYQKEMENDIQLYQSLEAKKMQVDLEKELQSAFNEITKLTSLIDGKVPKDLNLELEG 720
QY 721 KITDLOKELNKEVEENALREEVILLSELKSLPSEVERLRKEIQDKSEELHIITSEKDKL 780
DB 721 KITDLOKELNKEVEENALREEVILLSELKSLPSEVERLRKEIQDKSEELHIITSEKDKL 780

721 KITDLOKELNKEVEENALREEVILLSELKSLPSEVERLRKEIQDKSEELHIITSEKDKL 780
781 FSEVVHRESRVQGLLEIGIKTKDLDLATTOSSYKSTDOEFQNFKTLHDMDFQYKNVLEEN 840
781 FSEVVHRESRVQGLLEIGIKTKDLDLATTOSSYKSTDOEFQNFKTLHDMDFQYKNVLEEN 840
841 ERMQOEIVNLSKEAQKPFSSLGALKTSLSYKTOELQEKTRVQERLNEMQLEKQLENRD 900
841 ERMQOEIVNLSKEAQKPFSSLGALKTSLSYKTOELQEKTRVQERLNEMQLEKQLENRD 900
901 SPLQTVREKTLITTEKLOOTLEEVKTLTQEKDOLKQLOESIQIERDQKSDIHDITVNNVI 960
901 SPLQTVREKTLITTEKLOOTLEEVKTLTQEKDOLKQLOESIQIERDQKSDIHDITVNNVI 960
961 DTQEOLRNALSLKQHOFTINTLKSISEEVSRLNHEMTEGTDKDFQCKMVGIDKKQD 1020
961 DTQEOLRNALSLKQHOFTINTLKSISEEVSRLNHEMTEGTDKDFQCKMVGIDKKQD 1020
1021 LEAKNTQTLTADVKNENIEBOQRKIFSLIQEKNELOQMLSEVIAEKEQLKTDLKENIEMT 1080
1021 LEAKNTQTLTADVKNENIEBOQRKIFSLIQEKNELOQMLSEVIAEKEQLKTDLKENIEMT 1080
1081 IENQBELRLGDELKKQOEIVAEQKNHAIKKEGELSRCTCDRLAARVEEKLKESQOOLEKQ 1140
1081 IENQBELRLGDELKKQOEIVAEQKNHAIKKEGELSRCTCDRLAARVEEKLKESQOOLEKQ 1140
1141 QOLLNVQEMSEMOKINEIENLNKELNKLTLLEHMETERLELAQKLNENYEEVKSIITK 1200
1141 QOLLNVQEMSEMOKINEIENLNKELNKLTLLEHMETERLELAQKLNENYEEVKSIITK 1200
1201 ERKVLKEIQKSPETBRDLRGVIRIEATGLQTEBELKIAHILHKEHQFTIDELRRSVSE 1260
1201 ERKVLKEIQKSPETBRDLRGVIRIEATGLQTEBELKIAHILHKEHQFTIDELRRSVSE 1260
1261 KTAQIINTQDLEKSHTKLOEIPVLHBEQELLPNVKVSETOETMNELELLTEQSTTKDS 1320
1261 KTAQIINTQDLEKSHTKLOEIPVLHBEQELLPNVKVSETOETMNELELLTEQSTTKDS 1320
1321 TTARIEMERLRLNKEFQBSQOEIEXSLTKERONLAKITKEALFVKHDLKEHIRETLAKIQ 1380
1321 TTARIEMERLRLNKEFQBSQOEIEXSLTKERONLAKITKEALFVKHDLKEHIRETLAKIQ 1380
1381 ESQSQEQSLNKEKDNETTKIVSEMEQPKDQKDSALLRIEIMGLSKLOESHDKMSV 1440
1381 ESQSQEQSLNKEKDNETTKIVSEMEQPKDQKDSALLRIEIMGLSKLOESHDKMSV 1440
1441 AKEKDQLRLQELVQSESDQLKENIKEIVAKHLETEBELKVAHCCLKEQEBETINELRVNL 1500
1441 AKEKDQLRLQELVQSESDQLKENIKEIVAKHLETEBELKVAHCCLKEQEBETINELRVNL 1500
1501 SEKETEISTIQOLEAINDKLQNKIQEIVYEKEBOLNIKOISEVOENNELKQFKEHRKAK 1560
1501 SEKETEISTIQOLEAINDKLQNKIQEIVYEKEBOLNIKOISEVOENNELKQFKEHRKAK 1560
1561 DSALQIESKMLTNRLOESQOEIOMIKEEMKRVQEOALQIERDQKENTKEIVAKM 1620
1561 DSALQIESKMLTNRLOESQOEIOMIKEEMKRVQEOALQIERDQKENTKEIVAKM 1620
1621 KESQEKYQFLKMTAVNETQEKMCIEHLKQBPETOKLNLENTEINIRLTQILHENLEE 1680
1621 KESQEKYQFLKMTAVNETQEKMCIEHLKQBPETOKLNLENTEINIRLTQILHENLEE 1680
1681 MRSVTKERDRLRSVEETLKVVERDOLKENIRETITRDLKOEELKIVHMLKHEOETIDKL 1740
1681 MRSVTKERDRLRSVEETLKVVERDOLKENIRETITRDLKOEELKIVHMLKHEOETIDKL 1740
1741 RGIIVSEKTNESINMQKDEHSNDALKAQDLKIQELRIAHMLHKEQOETIDKRGIVSEK 1800
1741 RGIIVSEKTNESINMQKDEHSNDALKAQDLKIQELRIAHMLHKEQOETIDKRGIVSEK 1800
1801 TDKLSNNQKOLENSNAKLOEKIQELKANEHQLITLKQDVNETQKVSMEOLKKQIKDQS 1860
1801 TDKLSNNQKOLENSNAKLOEKIQELKANEHQLITLKQDVNETQKVSMEOLKKQIKDQS 1860

||||| 262 GSERAAQTGAAGVRLKEGGINRSLFILQVQIKKLSQVGGFINYRDSKLTIRLQNSLG 321
||||| 298 GNPKTRIICTITPVSFDETLTALQFASHTAKYKMKNTPPYNEVSTDEALLKRYRKEMDILKK 357
||||| 322 GNPKTRIICTITPVSFDETLTALQFASHTAKYKMKNTPPYNEVSTDEALLKRYRKEMDILKK 381
358 QLEVSLETRAQAEKQOQALQELLEKDLQVQNEKIENTRMLVTSSTLQOELKAKR 417
382 QLEVSLETRAQAEKQOQALQELLEKDLQVQNEKIENTRMLVTSSTLQOELKAKR 441
418 KRRVTWCGLGKINKMNSNYADQFNIPNTITTKTHKLSINLLREIDESVCSSESDVFSNTLD 477
442 KRRVTWCGLGKINKMNSNYADQFNIPNTITTKTHKLSINLLREIDESVCSSESDVFSNTLD 501
478 TLESEIENWPATKLINOENIESELNLRADYNLDVYEQLRTEKEEMELKLEKNDLDEF 537
502 TLESEIENWPATKLINOENIESELNLRADYNLDVYEQLRTEKEEMELKLEKNDLDEF 561
538 EALERKTKKQDBMOLIIHEISNLKLVHREVYNQDLENELSSKVLLREKEDQIKKLQY 597
562 EALERKTKKQDBMOLIIHEISNLKLVHREVYNQDLENELSSKVLLREKEDQIKKLQY 621
598 IDSQKLENIKMDLSYLSIESIEDPKQMKOTLFDATVVALDAKRESAFLSENLELKERKKE 657
622 IDSQKLENIKMDLSYLSIESIEDPKQMKOTLFDATVVALDAKRESAFLSENLELKERKKE 681
658 LATTYKQWENDIQLYQSOLEBAKKQWVDLEKLOSAFNEITKLTSLIDGKVPKOLLCNLE 717
682 LATTYKQWENDIQLYQSOLEBAKKQWVDLEKLOSAFNEITKLTSLIDGKVPKOLLCNLE 741
718 LEGKLTDLQKELNKEVEENEALREEVILLSELKSLPSEVERLRKEIQDKSEELHIITSEK 777
742 LEGKLTDLQKELNKEVEENEALREEVILLSELKSLPSEVERLRKEIQDKSEELHIITSEK 801
778 DKLFSEVVHKSERVQGLLEIEIGTKTKDDLAATTOSNYKSTDOEONFTLHMDPEQYKMWL 837
802 DKLFSEVVHKSERVQGLLEIEIGTKTKDDLAATTOSNYKSTDOEONFTLHMDPEQYKMWL 861
838 EENERNNOEIVNLSKEAQFDSLSGALKTEL SVKTOELKTEKREVOERLNEMBOLKEBLE 897
862 EENERNNOEIVNLSKEAQFDSLSGALKTEL SVKTOELKTEKREVOERLNEMBOLKEBLE 921
898 NRDSPLQTVVEREKTILITEKLOQTLEEVKTLTQEKDDLKOLQESLQIERDQKSDIHDVTN 957
922 NRDSPLQTVVEREKTILITEKLOQTLEEVKTLTQEKDDLKOLQESLQIERDQKSDIHDVTN 981
958 MNIDTQOELRNALLESUKQHOETINTLSKISSEVSRLNHEENTGETKDBFOQKMWGIDK 1017
982 MNIDTQOELRNALLESUKQHOETINTLSKISSEVSRLNHEENTGETKDBFOQKMWGIDK 1041
1018 KODLEAKNTOTLTADVKNONEIIEQQRKIFSLIQEKNELQOMLESVTAERKEQLTKDLKENI 1077
1042 KODLEAKNTOTLTADVKNONEIIEQQRKIFSLIQEKNELQOMLESVTAERKEQLTKDLKENI 1101
1078 EMTIENOEELRLIGDELKKQOEIVAQEKNHAIKKEGELSRTCDRLAEVEEKLKEKSQQLQ 1137
1102 EMTIENOEELRLIGDELKKQOEIVAQEKNHAIKKEGELSRTCDRLAEVEEKLKEKSQQLQ 1161
1138 EKQOQLLNVOERSEMOKKINEIENIKNELKNKELTLEHMETERLELAOKLNIENYEVKS 1197
1162 EKQOQLLNVOERSEMOKKINEIENIKNELKNKELTLEHMETERLELAOKLNIENYEVKS 1221
1198 ITKERKVLKELQSFETERDHLRGYIREIATGLQTKELKIAHILKEHOETIDELRRS 1257
1222 ITKERKVLKELQSFETERDHLRGYIREIATGLQTKELKIAHILKEHOETIDELRRS 1281
1258 VSEKTAQIINTQDLEKSHTKLOREIIPVLEHEQEALLPNVKVSVSETQMTMNELELLTQOSTT 1317
1282 VSEKTAQIINTQDLEKSHTKLOREIIPVLEHEQEALLPNVKVSVSETQMTMNELELLTQOSTT 1341
1318 KOSTTLARIEMERIRLNEKFOESQEI KSLTKERDNLKTIKBALEVKHQDLKEHIRETLA 1377
1342 KOSTTLARIEMERIRLNEKFOESQEI KSLTKERDNLKTIKBALEVKHQDLKEHIRETLA 1401
1378 KIQESQSQSQSQUAMKQKONETTQIVSEMEOFPKPKDSALLRIEIEMLGLSRLQESHDEM 1437
1402 KIQESQSQSQSQUAMKQKONETTQIVSEMEOFPKPKDSALLRIEIEMLGLSRLQESHDEM 1461
1438 KSVAKEXDILORLOEVLOSDDQKLENITVAKHLETEEBELKVAHCLCKEQUEETINELR 1497
1462 KSVAKEXDILORLOEVLOSDDQKLENITVAKHLETEEBELKVAHCLCKEQUEETINELR 1521
1498 VNLSEKETEISTIQKQLEAINDKLQKIQEIEYEKEQINIKOISEVQBNVNLQKFOKBEHR 1557
1522 VNLSEKETEISTIQKQLEAINDKLQKIQEIEYEKEQINIKOISEVQBNVNLQKFOKBEHR 1581
1558 KAKOSALQSIKSKMELTNRLOESQEBELQIMKEKEBMKRYOEALQIERDQKENTKEIV 1617
1582 KAKOSALQSIKSKMELTNRLOESQEBELQIMKEKEBMKRYOEALQIERDQKENTKEIV 1641
1618 AKMKESQEKYQFOLKMTAVNETQKMCIEHLKQFQETQKLNLENIENTENIRLTQILHEN 1677
1642 AKMKESQEKYQFOLKMTAVNETQKMCIEHLKQFQETQKLNLENIENTENIRLTQILHEN 1701
1678 LEEMRSVTKBRDDLSRVEETLKVVERDQKLENLRETITRDLKQEBELKIVHMHLEKHQETI 1737
1702 LEEMRSVTKBRDDLSRVEETLKVVERDQKLENLRETITRDLKQEBELKIVHMHLEKHQETI 1761
1738 DKLRGIVSEKNEISNMKOLLEHNSNDALKAQDLKIQEELRIAHMHLEKQOFTIDKLRGIV 1797
1762 DKLRGIVSEKNEISNMKOLLEHNSNDALKAQDLKIQEELRIAHMHLEKQOFTIDKLRGIV 1821
1798 SEKTDKLSNMKODLENSNAKLOEKIQELKANEHOLITLTKDVNETQKVSMEQOLKKQIK 1857
1822 SEKTDKLSNMKODLENSNAKLOEKIQELKANEHOLITLTKDVNETQKVSMEQOLKKQIK 1881
1858 DQSLTSLKLEIENLNAQELHENLEEMKSVKBERDNLRRVETLKLERDQKESLOFTKA 1917
1882 DQSLTSLKLEIENLNAQELHENLEEMKSVKBERDNLRRVETLKLERDQKESLOFTKA 1941
1918 RDLETOQELKTYARMLESKEHETVDKLREKISEKTIQISDIQKDLQSKDELQKQIOELQK 1977
1942 RDLETOQELKTYARMLESKEHETVDKLREKISEKTIQISDIQKDLQSKDELQKQIOELQK 2001
1978 KELQLLRVKEDVNMESHKKINEMEQKKOPEPNYLCCKCENDNFQTKLHESLEEIRIVAK 2037
2002 KELQLLRVKEDVNMESHKKINEMEQKKOPEPNYLCCKCENDNFQTKLHESLEEIRIVAK 2061
2038 ERDELRRIKESLKMEDQFIATLRMIARDRQNHQVPEKRLSDGQOQHLMESLREKCSR 2097
2062 ERDELRRIKESLKMEDQFIATLRMIARDRQNHQVPEKRLSDGQOQHLMESLREKCSR 2121
2098 IKELLKRYSEMDDHVECLNRLSLDLKEIETEFHRIKMKLVLSYVTYKIEEHECINKPE 2157
2122 IKELLKRYSEMDDHVECLNRLSLDLKEIETEFHRIKMKLVLSYVTYKIEEHECINKPE 2181
2158 MFDIPEVEKQKELLIKIQLHQDCDVPSELRLDLKLNQNMDLHIEILKDFSESEFPSTK 2217
2182 MFDIPEVEKQKELLIKIQLHQDCDVPSELRLDLKLNQNMDLHIEILKDFSESEFPSTK 2241
2218 TEFQOVLNKRKEMTQFLEEWLNTRFDIKLNKNGIQKENDRICOVNFFNRIAINVEST 2277
2242 TEFQOVLNKRKEMTQFLEEWLNTRFDIKLNKNGIQKENDRICOVNFFNRIAINVEST 2301
2278 EFERGATISKEWEQDLKSLKEKNEKLPKNYQTLTKTSLASGQVNTTQDNKNPHVTSRA 2337
2302 EFERGATISKEWEQDLKSLKEKNEKLPKNYQTLTKTSLASGQVNTTQDNKNPHVTSRA 2361
2338 TQUTTKIRLENSLHEAKESAMHESKIIKMQKELEVNDIITAKLQAKVHESNKCLEKT 2397
2362 TQUTTKIRLENSLHEAKESAMHESKIIKMQKELEVNDIITAKLQAKVHESNKCLEKT 2421
2398 KETIQVLQDKVALGAKPYKEEIEDLKNMVLKVIDLEKMKNAKEPEKEISATKATVEYQKEV 2457
2422 KETIQVLQDKVALGAKPYKEEIEDLKNMVLKVIDLEKMKNAKEPEKEISATKATVEYQKEV 2481

QY 2458 IRLLENLRRSQADQTSVISEHTDPPSNKPLTCGGSGIVQNTKALILKSEHIREKE 2517
 Db 2482 IRLLENLRRSQADQTSVISEHTDPPSNKPLTCGGSGIVQNTKALILKSEHIREKE 2541
 QY 2518 ISKLKQNEQLIKQKNEQLSNNQHLSENEVTKERTIKRAHKQVTCENSPKSPKVTGTA 2577
 Db 2542 ISKLKQNEQLIKQKNEQLSNNQHLSENEVTKERTIKRAHKQVTCENSPKSPKVTGTA 2601
 QY 2578 SKKKQITPSCQKERNLQDPVPKSPKSCFFDSRSKSLPSPHPVRYFDNSSIGLCPYONA 2637
 Db 2602 SKKKQITPSCQKERNLQDPVPKSPKSCFFDSRSKSLPSPHPVRYFDNSSIGLCPYONA 2661
 QY 2638 GABSVDSQPGPW-HASSGKDVPCECKTQ 2663
 Db 2662 GABSVDSQPGPWARLFQCKDVPCECKTQ 2688

RESULT 3
 ABG06505
 ID ABG06505 standard; protein; 2633 AA.
 XX
 AC ABG06505;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #6496.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US0008631.
 XX
 PR 31-MAR-2000; 2000US-00540217.
 PR 23-AUG-2000; 2000US-00649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR N-PSDB; AAS70692.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostic, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 PS Claim 20; SEQ ID NO 36864; 103pp; English.
 XX

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic

CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 2633 AA;
 Query Match 98.5%; Score 13131; DB 4; Length 2633;
 Best Local Similarity 99.4%; Pred. No. 0;
 Matches 2630; Conservative 0; Mismatches 2; Indels 14; Gaps 1;
 QY 1 MABEGAVAVCVVRPLNSREESLGETAQVYTKDNNVIYQVDSKSNFVRVHGNHTTK 60
 Db 1 MABEGAVAVCVVRPLNSREESLGETAQVYTKDNNVIYQVDSKSNFVRVHGNHTTK 60
 QY 61 NYEEETAAPIIDSAIOGYNGTIFAYGQTASGKTYTMMGSEHDHGVIPRAIHDFQIKKF 120
 Db 61 NYEEETAAPIIDSAIOGYNGTIFAYGQTASGKTYTMMGSEHDHGVIPRAIHDFQIKKF 120
 QY 121 PDREFLLRVSYMEIYNETITDLCGQTKMKPLIREDVNRVNVVADLTTEVVYTSEMAIK 180
 Db 121 PDREFLLRVSYMEIYNETITDLCGQTKMKPLIREDVNRVNVVADLTTEVVYTSEMAIK 180
 QY 181 WITKGEKSRHYGETKMNQSSRSHTIFRMILESRKGEPCNCGSVKSHNLVLDLAGSE 240
 Db 181 WITKGEKSRHYGETKMNQSSRSHTIFRMILESRKGEPCNCGSVKSHNLVLDLAGSE 240
 QY 241 RAAQTGAAGVRLKEGNCINRSFILQGVIKKLSGQVGVFINYRDSKLTILQNSLGGNP 300
 Db 241 RAAQTGAAGVRLKEGNCINRSFILQGVIKKLSGQVGVFINYRDSKLTILQNSLGGNP 300
 QY 301 KTRIICTITPVSFDETLTALQFASTAKYMKNTPYNNEVSTDEALLKRYKEIMDLKKQLE 360
 Db 301 KTRIICTITPVSFDETLTALQFASTAKYMKNTPYNNEVSTDEALLKRYKEIMDLKKQLE 360
 QY 361 EYSLETRAQAMEKQDLAQLLEKDLQKQNEKLENLRLMVLTSSTLTQOELKARKRR 420
 Db 361 EYSLETRAQAMEKQDLAQLLEKDLQKQNEKLENLRLMVLTSSTLTQOELKARKRR 420
 QY 421 VTWCLGKINMKNSYADQFNPTNITTKHKLINLREIDESVCSDESNTLDTLS 480
 Db 421 VTWCLGKINMKNSYADQFNPTNITTKHKLINLREIDESVCSDESNTLDTLS 480
 QY 481 ETEMNATKLNQENIESELNSIRADYDNLVDYQLRTEKEEMELKLEKNDLDFEAL 540
 Db 481 ETEMNATKLNQENIESELNSIRADYDNLVDYQLRTEKEEMELKLEKNDLDFEAL 540
 QY 541 ERKTKKQEMQLIHEISNLKLVKHEVYNQDLENELSKVELLEKEDQIKKQYIDS 600
 Db 541 ERKTKKQEMQLIHEISNLKLVKHEVYNQDLENELSKVELLEKEDQIKKQYIDS 600
 QY 601 QKLENIKMDLSYLSIESIEDPKQKOTLFDATVVALDAKRESAPLRSNLEKEMKELAT 660
 Db 601 QKLENIKMDLSYLSIESIEDPKQKOTLFDATVVALDAKRESAPLRSNLEKEMKELAT 660
 QY 661 TYKQENDIQYQSOLEAKKQVQDLEKELQSAFNEITKLSIDGKVPKDLCLNLEEG 720
 Db 661 TYKQENDIQYQSOLEAKKQVQDLEKELQSAFNEITKLSIDGKVPKDLCLNLEEG 720
 QY 721 KITDLQKELNKEVEANEALREEVILLSELKSLPSVERLRKEIQKSEELHIITSEKDL 780
 Db 721 KITDLQKELNKEVEANEALREEVILLSELKSLPSVERLRKEIQKSEELHIITSEKDL 780
 QY 781 FSEVVKHSRVQGLLEETGKTKNDLATTQSNYKSTQDFQNFPTLHMDPEQYKWLLEN 840
 Db 781 FSEVVKHSRVQGLLEETGKTKNDLATTQSNYKSTQDFQNFPTLHMDPEQYKWLLEN 840
 QY 841 ERMQEIVNLKSEAQKPFSSIGALKTELSTYKTOELOETREVOERLNEWEQKOLENRD 900
 Db 841 ERMQEIVNLKSEAQKPFSSIGALKTELSTYKTOELOETREVOERLNEWEQKOLENRD 900
 QY 901 SPLOTVEREKLITTEKLOOTLEEVKTLTOEKDLDLQKLESQIERDQKSDTHDNNNTI 960
 Db 901 SPLOTVEREKLITTEKLOOTLEEVKTLTOEKDLDLQKLESQIERDQKSDTHDNNNTI 960

Db 901 SPLQTVREKTLITEKLOQTLEEVKTLTQEKDDLKQLOESLQIRSDQKSDIHDTVNNMI 960
 Qy 961 DTQOLRNALSKQHOETTINTLKSKEEVRNLHMEENTGETKDFQOKWGDKKQD 1020
 Db 961 DTQOLRNALSKQHOETTINTLKSKEEVRNLHMEENTGETKDFQOKWGDKKQD 1020
 Qy 1021 LEAKNTQTLTADVDNDNEIEEQRKIFSLIOEKNELOQMLESVIAEKQKLTDLKENIEMT 1080
 Db 1021 LEAKNTQTLTADVDNEIEEQRKIFSLIOEKNELOQMLESVIAEKQKLTDLKENIEMT 1080
 Qy 1081 IENQERLRLGDELKKQOEIVAOEKNHAIKEGELSRCTDLAVVEKLEKSKSOLOEQK 1140
 Db 1081 IENQERLRLGDELKKQOEIVAOEKNHAIKEGELSRCTDLAVVEKLEKSKSOLOEQK 1140
 Qy 1141 QOLLNVQESMOKKINEENLNELNKLKELTLEHMETRELELAQKLNENYEVKSITK 1200
 Db 1141 QOLLNVQESMOKKINEENLNELNKLKELTLEHMETRELELAQKLNENYEVKSITK 1200
 Qy 1201 ERKVKELQKSFETERDHLRGYIREIEATGLQTKBELKIAHILKEHQETIDELRRSVSE 1260
 Db 1201 ERKVKELQKSFETERDHLRGYIREIEATGLQTKBELKIAHILKEHQETIDELRRSVSE 1260
 Qy 1261 KTAQIINTQDLKESHTKLOEIPVLHBEQELLPNVKVSETQETWNELELLTEOSTTKDS 1320
 Db 1261 KTAQIINTQDLKESHTKLOEIPVLHBEQELLPNVKVSETQETWNELELLTEOSTTKDS 1320
 Qy 1321 TTLARIEMERLRLNEKFOESQEEIKSLTKERDNLKTIKEALEVKHDQKHEIRETLAKIQ 1380
 Db 1321 TTLARIEMERLRLNEKFOESQEEIKSLTKERDNLKTIKEALEVKHDQKHEIRETLAKIQ 1380
 Qy 1381 ESQSKQESQSLNMEKNDNTTIVSEMEQFKPDSALLRIEIMGLSKRLQESHDEMKSV 1440
 Db 1381 ESQSKQESQSLNMEKNDNTTIVSEMEQFKPDSALLRIEIMGLSKRLQESHDEMKSV 1440
 Qy 1441 AKEDDLQRLQEVLOSQDQKKEIYVAKHLETEBELKVAHCCLEKEQETINEELRVNL 1500
 Db 1441 AKEDDLQRLQEVLOSQDQKKEIYVAKHLETEBELKVAHCCLEKEQETINEELRVNL 1500
 Qy 1501 SEKETEISTIOKQLEAINDKQNKIOEIVEEQLNIOISVQENVNVELKQFKEHRKAK 1560
 Db 1501 SEKETEISTIOKQLEAINDKQNKIOEIVEEQLNIOISVQENVNVELKQFKEHRKAK 1560
 Qy 1561 DSALQIESKMLNLRNLSQEEIOMIKEEMKRVQEAQIQRDQKENTKEIVAMK 1620
 Db 1561 DSALQIESKMLNLRNLSQEEIOMIKEEMKRVQEAQIQRDQKENTKEIVAMK 1620
 Qy 1621 KESQKEYOFLQMTAVNETQKMCIEIHLKEQFETQKLNLENTENTELTQILHENLEE 1680
 Db 1621 KESQKEYOFLQMTAVNETQKMCIEIHLKEQFETQKLNLENTENTELTQILHENLEE 1680
 Qy 1681 MRSVTKERDDLSRVETLKVVERDQKLENLREITITRDLEKQEBELKIVHMLKEHQETIDKL 1740
 Db 1681 MRSVTKERDDLSRVETLKVVERDQKLENLREITITRDLEKQEBELKIVHMLKEHQETIDKL 1740
 Qy 1741 RGIVSEKTEINISNMQKDLHSNDALKAQDLKTOEELRIAHMHLKEQOETIDKLGIVSEK 1800
 Db 1741 RGIVSEKTEINISNMQKDLHSNDALKAQDLKTOEELRIAHMHLKEQOETIDKLGIVSEK 1800
 Qy 1801 TDKLSNMQKDLHSNAKLOEKIOELKANEHOLITLKKOVNETQKVSMEQLKKQIKQDS 1860
 Db 1801 TDKLSNMQKDLHSNAKLOEKIOELKANEHOLITLKKOVNETQKVSMEQLKKQIKQDS 1860
 Qy 1861 LTLKLEINLNLAQELHENLEEMKSVKMERDNLRRVEETLKLERDQKESLQETKARDL 1920
 Db 1861 LTLKLEINLNLAQELHENLEEMKSVKMERDNLRRVEETLKLERDQKESLQETKARDL 1920
 Qy 1921 ETQOELKTARMLSKEHETVTKLREKISEKTIQISDIQKDLKSKDELQKLOKKELOK 1980
 Db 1921 ETQOELKTC-----SEKISEKTIQISDIQKDLKSKDELQKLOKKELOK 1966
 Qy 1981 QLLRVKEDVNMHSHKKNEMEQKQFEPNVLCKCENDNFQTKKLHESLEETIRIVAKERD 2040
 Db 1967 QLLRVKEDVNMHSHKKNEMEQKQFEPNVLCKCENDNFQTKKLHESLEETIRIVAKERD 2026

Qy 2041 ELRRIKESLQVBERDOFIATIREMTARDQRONHVKPEKRLSDGQOHLMESLREKCSRIKE 2100
 Db 2027 ELRRIKESLQVBERDOFIATIREMTARDQRONHVKPEKRLSDGQOHLMESLREKCSRIKE 2086
 Qy 2101 LLKRYSEMDDHYECNLNLSLDLEKEIEFHRIMKKLVLSVVTYKIEQEHCEKCNKFMDF 2160
 Db 2087 LLKRYSEMDDHYECNLNLSLDLEKEIEFHRIMKKLVLSVVTYKIEQEHCEKCNKFMDF 2146
 Qy 2161 IDEVEKQKELLIKIQLHQQDCDVPSERLDLKLNQNDLHIEELKDFSESEFFPSIKTEF 2220
 Db 2147 IDEVEKQKELLIKIQLHQQDCDVPSERLDLKLNQNDLHIEELKDFSESEFFPSIKTEF 2206
 Qy 2221 QQVLSNRKEMTOFLEEWLNTRFDIEKLKNGIOKENDRICQVNNFNFRILIAIMNESETEFE 2280
 Db 2207 QQVLSNRKEMTOFLEEWLNTRFDIEKLKNGIOKENDRICQVNNFNFRILIAIMNESETEFE 2266
 Qy 2281 ERSATISKWBODLKSLEKNEKLFKNVQTLKTSLSAGQAVNPTTQDNKNPHVTSRAEQ 2340
 Db 2267 ERSATISKWBODLKSLEKNEKLFKNVQTLKTSLSAGQAVNPTTQDNKNPHVTSRAEQ 2326
 Qy 2341 TTEKIRELENSLHSAKSSAMHESKIIKMOKELEVTNDIIAKLQAKVHESNCKLEKTKET 2400
 Db 2327 TTEKIRELENSLHSAKSSAMHESKIIKMOKELEVTNDIIAKLQAKVHESNCKLEKTKET 2386
 Qy 2401 IQVLDQKVALGAKPYKEIEIDLKMKLVKIDLEKMKNAKEPEKISATKATVYQKEVIRL 2460
 Db 2387 IQVLDQKVALGAKPYKEIEIDLKMKLVKIDLEKMKNAKEPEKISATKATVYQKEVIRL 2446
 Qy 2461 LRENLRSSQQAQDTSVISEHTDPOPSNKPITCGGSGGIVQNTKALILKSEHIREKEISK 2520
 Db 2447 LRENLRSSQQAQDTSVISEHTDPOPSNKPITCGGSGGIVQNTKALILKSEHIREKEISK 2506
 Qy 2521 LKQNEQLIKQKNELLSNNHLSNEVTKWERTLKREAHKQVTCENSFKSPKVTGTASKK 2580
 Db 2507 LKQNEQLIKQKNELLSNNHLSNEVTKWERTLKREAHKQVTCENSFKSPKVTGTASKK 2566
 Qy 2581 KQITPSQCKERNLODPVPKESPKSCFFDSRSKSLPSHPVRYFNDSSILGLCPEVQNAAGAE 2640
 Db 2567 KQITPSQCKERNLODPVPKESPKSCFFDSRSKSLPSHPVRYFNDSSILGLCPEVQNAAGAE 2626
 Qy 2641 SVDSQP 2646
 Db 2627 SVDSQP 2632
 RESULT 4
 AA01632
 ID AA01632 standard; protein; 2954 AA.
 AC AA01632;
 DT 22-JUN-1999 (first entry)
 DE Amino acid sequence of centromere-associated protein-E (CENP-E).
 KW CENP-E; centromere-associated protein-E; ATPase activity;
 KW plus end-directed microtubule motor activity; chromosome congression;
 KW microtubule binding activity; chromosome movement; mitosis;
 KW cell proliferation; tumor; metastasis; vascular malfunction;
 KW inflammatory disease; immune disease; angiogenesis; hypertension;
 KW restenosis; fungal infection; selective herbicide; fungicide;
 KW insecticide; plant growth regulator; activator; cancer cell marker.
 OS Xenopus sp.
 PN WO9913061-A1.
 XX 18-MAR-1999.
 PD 10-SEP-1998; 98WO-US019231.
 PF 11-SEP-1997; 97US-0058645P.
 PR

Db 1618 LLEELHOKNEQKLAHNELEQAQVELKCEVEHLMKMSIESKLSLESQHEKHDTQ 1677
Qy 1478 ELKVAHCLKEQEBETINELVNLSKE-----TELSTIQKLEANDKLNKIQIYEKEE 1533
Db 1678 QL-----LALKQOMQVVTQEKELQOTHEHLTAFAVDHLKENITE-LGLNFKNEAQKTKKEQ 1732
Qy 1534 QL--NIKOISVQENVN-ELKOFKHEHRAKADSAQSI-----ESKMLELTNRLQESQETQI 1587
Db 1733 CLINENKELESQURLQCEIEELMSKLDKESALETLKESQKVINL-----NQMEM 1785
Qy 1588 MIKEKEEMKRYQEAQIOLQERDQKLENTKEIIVAMKESQF--KEYQFLKMTAVNETQKCE 1645
Db 1786 VMEWEELKNSQRTVIAERDQLODDLRSEVMSIETQDDLRKAQEAQLOQKQKQVQELTSQ 1845
Qy 1646 IEHLKEQEPQKMLNENITENIRLTQILHENLEEMSVTKERDDLRSEVETLKVVERDOL 1705
Db 1846 ISVLQEKISL-----LEN-----QMLY-NVATVKTETLSEERDNLQSKHLSIEITL 1891
Qy 1706 KENRETTITROLEKQELKIVHMLKEHQETID---KLRGIVSEKTNIEISNMOKDLEHSN 1762
Db 1892 SLSUKEK-EPALQEAQKDK-----ADAARKIIDITEKISNIEEQLLOQATNLKETL-YER 1944
Qy 1763 DALKAQDQKIQEBRLAHMHLEQOETIDKLRGIVSEKTDKLSNMOKDLENSAKLQEKI 1822
Db 1945 ESL-----IQCKEQLALNTEHLRETLKSKDALGKMEQERDEAANKVIALTEKMSLEBQI 2000
Qy 1823 QE-----LXANEHOLITLKKDVNETQKVSMEOLAKQIKDQSLTSLKL-----EIE 1869
Db 2001 NENVTLKEGEGEKETFYLRPSKQOSSQMEELRESLTKDQLEBAEKEISEATNEIK 2060
Qy 1870 NL-----NLAQELHENLEMSVMKERNLNRVEETLKLERDQKESLQETKARDLEIQ 1924
Db 2061 NLTKISSLEBEILQNASILNEANVERENLRHSKQQLVSEQLSLTL---KSRDHAFQA 2117
Qy 1925 ELKTARMLSKHEKTVDKLREKISSEKTIQISDIQKDLK---SKBELQKKIQELO----- 1976
Db 2118 -----SKREK---DEAVNKIASIABEIKILTKEMDFRDSKESLQESQSHLSSEELCT 2166
Qy 1977 -KKELQALLR-VKEDVMS-HKKINEMEQKKQFEPNYLCKCEMDNFOLTKLHESLEIR 2033
Db 2167 YKTELQMLKQKQEDINNLAKEKVEDEL-----LOHLSLSKEQLDQIQ 2210
Qy 2034 IVAKERDELRLKESLQME-RDQFIATREMIARDQRNHQVPEKRLSLSDGOQHLMESLR 2092
Db 2211 MELR-NEKLRNVELCEKMDIMEKEISVLRM-----QNEPQOE-EDDVAERMDILE 2259
Qy 2093 EKCSRIKELLKRYSEM-DHYECLNRLSLDLKEKEIEFHR----- 2130
Db 2260 SRNQEIQELMEKISAVYSQHTLLSSLSSELOKETEHAHKMLNIKESLSSTLSRSGSL 2319
Qy 2131 -----IMKKLAVLSYVTKIEEQHECINKPEMDPIDVEKOKELLIKIOHL 2177
Db 2320 QTEHVKNLTQLOTLNKFV--YRTAAVKEDHSLIKDYEKDLAAQKHDELRLQLOQL 2377
Qy 2178 QOD-----CDVPSRELADKLNMOMDL-----HIEBILKDFSESE--FPSIKTEFOOVILS 2225
Db 2378 EQHGRKWSDSASELAKFCIEFLNELLFKKANIQSVQDDFSEVQVFNQVQSTLQEELE 2437
Qy 2226 NRKEMTQFLEEMINTFDEKLNKNGQKENDRICQVNNFNPNRIIIMNESTEFERSAT 2285
Db 2438 HKKGFMQWLEEFGLDHVDAKLKSEGQENRRIASTIQLTKRLKAVQSKI---QREIT 2494
Qy 2286 I-SKEWEQDLKSLKEKNEKLFKNYOTLTKTSLASGAOVN-----PTQONKNPHVTSRA 2337
Db 2495 VYINQPEAKLQEKQKQNKELMRMEHHGPGASVMEENARLLGILKTVQDE-----SKX 2548
Qy 2338 TQLTTBKIRELSLHSAKESAMHESKSIIMQKQELVTDIITAKL---QAKVHESNKCL 2394
Db 2549 LQ---SRIKMLENELNLVKDDAMHKGKVAILOKQ-LSRNABAEALNMQVLTQKQDNL 2604
Qy 2395 EKYKETIOVLQDKVALGAKYKEIEDLKMKLKIDLEKMKNAKEFEKESATKATVEYQ 2454

Db 2605 QAAMKEIENLQKVAKGAVPYKEEIDNLTKVVKIEMEKIKYSKATQEIAYLKSCLQEDK 2664
Qy 2455 KEVIRLLREMLRSQQADTSVISEHTDPOPSNKPPLTCGGSGGIVQNTKALILKSEHRL 2514
Db 2665 EELRLRLKEELRAQADNDITVCVPKDYQKASTFPVTCGGSGGIVQNTKALILKSEHRL 2724
Qy 2515 EKEISKILKQNEQL-----IKONELLSSNOHLSNEVKTWKERTLKREAHKQV 2562
Db 2725 ERELSHYKKYVHLSRTMSSSEDRKTKAKSDAHSSTGSHRSPHKTYR---HGPV 2781
Qy 2563 TCE-----NSPK-----SPKVTGT-----ASK 2579
Db 2782 TPERSEMPSLHLGSPKSESSTKRVSPNRSEIYQLVMSPGTKGMHKHILSPKVGHLK 2841
Qy 2580 KQKITP-----SCKERNLQDPVPKESPKSCFF 2607
Db 2842 KRALSPNRSEMPTOHVISPGKTIHLKNTLSTFLDNLSSPCKQKQVQENL--NSPKGLKLF 2899
Qy 2608 DSRKSLSPHPVYFQNSISGLCPVQNAQESVDSQGPWHASSGKDVPECKT 2662
Db 2900 DVKSKSPY-CPSQFFDNKGLDFSELNTAESNDKSAENWWYBAKETAPECKT 2953

RESULT 5
ABB62322
ID ABB62322 standard; protein; 2013 AA.
XX
AC ABB62322;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 13758.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
PD 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
XX
XX 11-JUL-2000; 2000US-00614150.
XX
XX (PEXE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
XX N-PSDB; ABL06425.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions.
XX
XX Disclosure; SEQ ID NO 13758; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceuticals. The invention
XX discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
XX sequences (AB101840-AB116175) and the encoded proteins (AB57737-
XX AB57702). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 2013 AA;

Query Match	11.6%;	Score 1545.5;	DB 4;	Length 2013;	
Best Local Similarity	25.9%;	Pred. No. 2e-58;			
Matches	590;	Conservative 425;	Mismatches 752;	Indels 515;	Gaps 92;
QY	6	AVAVCVVRPLNSREESLGTAYWYKT-DNNVIYQVDG-SKSNFVRVFGHNETTKNYY	63		
Db	8	SIQCIKVRCEPGLTSL-----WQVKERRSIHLADSHAEPYVYVDFVDEGASNOEVF	60		
QY	64	BEIAAPIIDSAIQYNGTIFAYGOTASGKTVMGSDHLGVI-PRATHIDFQIKKFPDR	123		
Db	61	DRMAKHIVHACMGFGNGTIFAYGOTSGKTYTMGDQONPGVWVLAKEIPQQLSSETER	120		
QY	124	EFLRVSYMIYNETITDILCGTQMKPLI-IREVDNRNVYVADITEEVYVYSEM-ALKWI	182		
Db	121	DFLRVGYIETNEKIVYDIL--NKKNQDLKTHESGNGIVNVN--CEECIIITSEVDLRL	176		
QY	183	TKGKSRHYGCTKQNRQSSSHITFPMILSREKGFSPNCEGSKVSHLNLVDLAGSRA	242		
Db	177	CLGNKERTVGETNWRSSSRSHAFKIIIESR-KSDHSD-DDAVIQSVLNLVDLAGSRA	234		
QY	243	AQTGAAGVRLKEGCNINRSFILGOVTKLSDGVQGVFINVRDSKLTIRILQNSIGNPKT	302		
Db	235	DQTGARGARKGHHINKSLPFLSNVIXLSENADNFTWYRDSKLTIRILQASLGNAFT	294		
QY	303	RIICTTPVSDFTLALOPASTAKYMKNTYPYNEVSTDEALKRYRKEIMDKKOLEEV	362		
Db	295	SIICTIPSIEMBSQSTLSPATRAKKIRIKPQVNMVSDATMKRLEREIKVLKDLAE-	353		
QY	363	SLETRAQAMEKQDLAQLLEKDLQKQVNEKINLFTM-----LVTSSSLTLOQLKA	415		
Db	354	-----EER-----KXNQKVEHLERQIKHDMHKIICGHSLS-----DKG	388		
QY	416	KRRRRVTWC-----LGKINKMKNYADQF-----NIPNITTKTHKL-----SI	455		
Db	389	QKRRRTWCTPAGSHLELAETGTTEDRIDQPKVSHLPKPVFFHTSNACKRWNDIPKTI	448		
QY	456	NILREIDESVCSVSFNTDLTSLBIENWPATKLINQENIESELNSLRADYDNLVDYE	515		
Db	449	NILGSLD--IGTBS-----NSISE-EFLPAECI-----DFGSPRPDLVKPMLTIR	490		
QY	516	QARTEKEEMELKLEKNDLDEPALERKTKQDEMQLIHEISNLKVLKHREYVNOQLEN	575		
Db	491	QL-----PDLPTPQGPL-----TTDKTKKE-----IDQLMFTSLEKHFEVECEVQ	534		
QY	576	ELSSKVELLEREKEDQIKGLQYIDSKLENKIMDLVSLESIBDPKMKOTLFDATVAL	635		
Db	535	-----LKEKLAETVARDNLQESL-----AKERYDALEKEVTSU	570		
QY	636	DAKRESAFIRSENLEKKEKELATTYKQWENDIQL---YQSQLEAKK---KMQVD----	685		
Db	571	RADNEAA--NSKISELEELKSLTKQTMRTIMEVENQVAVGLEPEFAHKKSSKLVRVDDL	628		
QY	686	--LEKE-----LOSANETLK--LTSLDIG-----KVPKDLLCML--ELEKIID	724		
Db	629	ALLEKESTIESQKSLDNTRVLRNSKBGMLSTIAPEQEDTAGDSICNCKEELKLIAD	688		
QY	725	LOKELNKEVEENEALREEVILLSELKSLPSEVERLRKEIQDKSEELHIITSEKDKLFSV	784		
Db	689	LESKNS-----CECQDLREIVSRDKLESVESAFNLASSEI	726		
QY	785	VHKBRVQGLLEIGTKTDLDLATTQSNYSTDQEFQ-----NFKTLHMDPEQYKMWLEEN	840		
Db	727	IQKATDCERLSKELSTLSQAFQQLERYDALDQWQAQAGITTLH-----EKHEHVQEKY	782		
QY	841	ERMNQEVNLSKEAQKFDSDSIALKTELSTYKQELQEKTRVQERLNEMEQLEQLEND	900		
Db	783	QKQBEYEOLRESAR---SASABFORLQNDNTKFQADIASLNERLEEAQNMLTEVQNSE	839		
QY	901	SPLQTVEREKTLITELKQOTLEEVKTLTQEKDDL-KQLOESLQIERDQLSKDIHDTVMN	959		
Db	840	STVEKLRIQNHLETKIKELETFNEMQREYDCLSNQLMSESQVQ-ENDALREIEKORPTSH	898		
QY	960	IDTQSLRNALESLSKHQOBTINTLKS--KISEVSRNLHMEENTGETKDEPQKMWGIDK	1017		
Db	899	VEBSMRSGSISSDFDEQKQDINLLHQFVQLSESV-----	932		
QY	1018	KQLEAKNTQTLTADVXNDNIIIFQQRKIFSLIQEKNELOQMLESVIAEKEQLTKDLKENI	1077		
Db	933	-QOIELQHSGISRLFRANQW-----KLDQSEPGKLCLESA-----EYIEDNRQS-	978		
QY	1078	EMTIENQOELRLILGDELKXQOEIVAQEKNAHKKGBELSRTCDBLAEEVEKLEKKSQOLQ	1137		
Db	979	-----DATEPICLGFGLKRHRFQIKRLSDEHV--DMGEKRLDIIISQLEOIEEKESALME	1032		
QY	1138	EKQQQLNVQEBEMCKKINEIKNELKNKELTLEHMETERLELAQKLNENYEVKS	1197		
Db	1033	ATATINEMEQTNLESALLEKSVINKVEDVQRIESEKQNAEMTV-----YESLOD	1088		
QY	1198	-ITKERVKLKELQKSPETERDHLRG-----YIREIATGLQTKELKIAHILKHEQETI	1251		
Db	1089	RVTRESSMSLLR-VPPDEDTLPGCPTSPSRREQEVATLKT-----	1131		
QY	1252	DELRRSVSEKTAQIINTQDLEKSHTKLOEIPVLHREOELLP--NVKVSETOETMNELE	1309		
Db	1132	TELQSOVSQDLKABLEN-----HLRQIQDKGNTARLQTDPEMSERC	1173		
QY	1310	LITEQSTTKDSTTLARIEMERLRLNEKFOESQEEI-KSLTKERDNLKTIKEALEVHKDOL	1368		
Db	1174	L-----SWEVRLAELDEDTKQBELDRQAQKLSDDLCLI--DQIQKNAQL	1218		
QY	1369	KEHIRETLAKIQBSQSQBSQSLNMKKNBTTKIVSEMEQFKP-----KDSALLRIE	1420		
Db	1219	VEQYKATESLSLADAKPQILLSSQVDSQIEKLNQLLNAAKDBLHDVRRIKDDEISALR	1278		
QY	1421	IEMGLSKRLOESHDEKSVAKEDDLQRLQEVLOQESDOLKENIKIIVAKHLEBEEL-	1479		
Db	1279	MEPL-----LOIETNEKENQAFYAELOETKDRYESNVAELEKLIQVEETLS	1326		
QY	1480	KVAHCCLKEQETINELRVNLSEKETEISTI-----OKQLEAINDKLQNKIQEI---YEKE	1532		
Db	1327	SVTVRCQAELEALKSAHKENISOAVEERNLIVQOAEEMETTRKLNKLEASTQOOSKM	1386		
QY	1533	EOLNIKQISEVQENVNE-LKQFKHRAKADSAQSTESKMLELTN-----RLQES	1581		
Db	1387	EDAFRAEINEVRATLMQNLQTKEDRDKGASKLEEVKKTLEQMGINGRVMSDITAELEKT	1446		
QY	1582	OEEIQIMIKE-----KEEMKRVQEAQOIE---RDQLK-----ENTKEIVAKWK--	1621		
Db	1447	KAQDLAVNKLTKDNIELEKQCKTQEQLOMESLTDQISFIEAHIKKLELIVASSKXR	1506		
QY	1622	--ESQEK-EYQFL-----KMTAVNETQ-----EKMCIEHLKE-----QFETQ	1656		
Db	1507	IIELEKCDQOVLELDKCELEKLSLESELOKANSEHSCMTWEKLOELQAEKVLNREKE	1566		
QY	1657	KUNLE-NIETENIRLTQILHENLEEMRSVTKPRDDLRSVEETLKV---ERDQIKENLRET	1712		
Db	1567	KCDFETKLETFTFKITD-LEEVLKEAQHKVILYDDLVSOHERLKI CLAEANELSSNLQKK	1625		
QY	1713	ITFDLEKQELKIVHMLKHEQETIDKLRGIVSEKTEINISNMQKOLEHENDA---LKAQD	1769		
Db	1626	V-----MSLTHELIDSQGI--SSRDVEINELREELKAAMDAKATASAEQ	1668		
QY	1770	LKIQBELRIAHMHLKEQOETIDK---LRGIVSEKTDKLSNMQ--KD-LENSNAKLOEKI	1822		
Db	1669	MTLVTLQKDVEERMANQAEKFTREANLKGSLNELLKNSMQETKMDLESNEELKEQL	1728		
QY	1823	QB-----LKANEHOLITLKQVNETQKVSMEBQL-----KKQIKDOSLTSLKLEIE-N	1870		
Db	1729	RNSQNLNMLDESKWCISLKEKIVLEDAKTSLSQQLRDKNKSEIYQRIHETLK-EVELG	1787		
QY	1871	LMLAQELHENLEEMKSVMKERDNLRVBETLKLERDQKLESQETKARDLETOQELKTAR	1930		
Db	1788	RNRIGELTKKCBELCSLSENSQIR-----LDIQETKEQLKKTLENNLQWQ-----	1834		
QY	1931	MLSKHEKHTVDKLRKIKSEKTIQISDIQKDLKSKDELQKQLOELQKELQLLRVKEDVN	1990		

Db 1835 -----KVDEVTRECEKLRFDWQSKVEQVSKVQEL----- 1864
 QY 1991 MSHKKINEMELQKQFENYLCKEMDNFOLTKKLHESLEIRIVAKERDELRIKESL- 2049
 Db 1865 -----ISECEBELSTL-----KSKASFOSEK-----ESMD--RISSILLEKRNLEKLC 1908
 QY 2050 -----KMERDQFIATIREMIARDRQ-----NHQVKEKRL--SDGQOHLMESL 2091
 Db 1909 SANDIVAKLEFE--IAALRPKSLDRNPVPRKSTTFESEIRKRRISVHDERROSYNDV 1966
 QY 2092 RE 2093
 Db 1967 RE 1968
 RESULT 6
 ID ABB61012 standard; protein; 1931 AA.
 XX ABB61012;
 AC ABB61012;
 XX 26-MAR-2002 (first entry)
 DT Drosophila melanogaster polypeptide SEQ ID NO 9828.
 DE Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX Drosophila melanogaster.
 OS WO200171042-A2.
 PN 27-SEP-2001.
 PD 23-MAR-2001; 2001WO-US009231.
 PF 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX (PEKE) PE CORP NY.
 PA Venter JC, Adams M, Li PWD, Myers BW;
 PI WPI; 2001-656860/75.
 DR N-PSDB; ABL05115.
 DR New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.
 PS Disclosure; SEQ ID NO 9828; 21pp + Sequence Listing; English.
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signaling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX Sequence 1931 AA;
 SQ
 Query Match 9.6%; Score 1286; DB 4; Length 1931;
 Best Local Similarity 23.8%; Pred. No. 3.2e-47;
 Matches 527; Conservative 418; Mismatches 767; Indels 500; Gaps 84;
 QY 6 AVACVVRPLNSREESLGTAQVYWKTDNNVIYQVDGS--KSFNEDRVFPHGNETTKNVY 63
 Db 8 SIQVCIKRCPCEGLTSL-----WQVKEGRSIQLADSHAPYVVDYVDEGASNQEVF 60

QY 64 BEIAPIIDSAIOGYNGCTIFAYGOTASGKTYTMMGSEDLGLVIPAIAHDIPOKIKFPDR 123
 Db 61 DRMAKHIVHACMGQNSGTIFAYGQTSIGSKTYTMMGDEONFGVMVLAKEIFQOISSETER 120
 QY 124 EFLLRVSMELYNETITDLCGTOKMKPLIREDVNRNVYVADLTEEVYVTSMALKWIT 183
 Db 121 DFLRVGYIHYNEKIYDYL--NKNQDLKHESGNGIVNV--NCKESIVTSEDDLLRQLY 177
 QY 184 KGEKSRHYGETKMQRSRSHHTIPRMILESEKPEPCSGSVKSHNLVDLAGSRAA 243
 Db 178 MGNERVVGETNMNERSRSHAIPIRIIESKPSHDN--DTVKOSVLSVDLAGSEQVD 235
 QY 244 QTGAAGVRLKEGCNINESLFILOVKKLSGQGVGFNYRDSKLTIRLQNSLGNPKTR 303
 Db 236 PADHAS-----SLMIFRNLVKSLSVDSKPNFSFDSKLPRIMLPSLGGNVLTS 284
 QY 304 IICITPVSFDETLTALQFASTAKYMKNTPTVNEVSTDEALLKRYKEIMDLKKQLEVS 363
 Db 285 IICITPVSFVESSSTISFGTCAKKIRCKPQCKIDSETTMW-----RQLD--- 330
 QY 364 LETRAQAMEKDOLAQLLEEKOLLQKQVNEKIENITRMVLVTSSSLTLOQELAKAKRRVTM 423
 Db 331 ---RGISMLKDKLAK-----KKIKNE-----SOLVL-QELEGRIKRDML- 365
 QY 424 CLGKINKMNSNYADQFNIPNTITTKHLINLREIDESVCSGESDVFSNTDLSIE 483
 Db 366 -----KIVSSASLDDYRLQKRRTWLTAS-----GSEGDAPVALPEPESR 408
 QY 484 WNPATKLNQENIESELSRADYDNLVLYEQLTEKEEMELKLEKNDLDEFALERK 543
 Db 409 LPRPSKLTNPKPLFQRRGIAPKAGICKTLKEKRLQTDNMD----- 450
 QY 544 TKDOEMOLIHEISNLKMLVHREYVNOLEN-----ELSKV-----ELLREKEDQ 590
 Db 451 TMPGRAKQLGRETSRIEPSVMMSKQYQSVFNCDAPOTEISALTASNOVAKETIEKYE 510
 QY 591 IKKQLEYIDSOKLENIKMDLSVLESIEDPKMQKTLFDAETVALDRAKRESAFLSENLE 650
 Db 511 VARUKETIERLEMENGA-----VN 530
 QY 651 LKERMKELATTYKOMENDIOLYQSOLEAKKQVQVDEKLEKQAFNEITKLTSLIDGKVPK 710
 Db 531 LGEQFETHKAKSQMEEL--LSSISEKOSTIVSLQOSLE-----ELSR 572
 QY 711 DLLCNLELBGKITDLOKELNKEVER--NEALREEVIL--LSELKSLPSEVERLRKEIQ 764
 Db 573 DYLRNSK-EDQWRSKMCPELESSCERICNKCLELERLLPLASASGLDSVACQDFQLRSEIA 631
 QY 765 DKSEELHIITSEKDKLFSEVVHKEKRVQGLLEIGKTKDDLTATQSNYKSTDOEFQNFK- 823
 Db 632 ATRMKLESMLSTFSSHASCEVSQKTTDCRKLSEIQTSTAHDDFQQLQEKYNNLKHKWSQKL 691
 QY 824 ---TLHMDP---EQYKVMVLENE-----RMNQEI NVLSKAEKFOSSGALKTSLKYKQ 873
 Db 692 AIDTWQVDYNTIQKYQLQDEYRHLERLSDQCOQLQDENSCKLOAEIGTLKE-----RVE 747
 QY 874 ELQEKTRV---QERLNEHQLEQLENRDSPLQ-----TVREKTLITKELQOTLEEVKT 926
 Db 748 EIHSELLEVPNDTHPEDMELQNLQKRLSKLQWFEFQIQLNYECLSELMSLTQECDA 807
 QY 927 LTQE-----KDDLQOLESLOIERDLKSDTHDVTN--MNIDTQEQLRNALSLKQ----- 975
 Db 808 LREEHKQRTTNSDLESMSKSS-----GVGTSCSDPENELDLDLQQTFLSKSIQIILT 861
 QY 976 -----HOETINTLKS-KISEVSRNLHMEENTGETKDEFOQKMWGIDKKQDLE 1022
 Db 862 DYSGGRRFLFYNHAEQDQVPSLKLCLPEAKYL---EGDGKQHDASDSVFL---KGFLK 914
 QY 1023 AKNTQTLTADVKDNEIIEQORK---IFSLIOEKNELOQMLSVIAKEQLTKDLKENIEM 1079
 Db 915 CORFOIVKINEQNLVKEEDMRDIIIFOLKQEVGDKKNLIE-----BEKEVINNLRAQITS 970
 1080 -----TIENQOE-ELRLIGDELKKQEIIVA---QEKHAIKKEGELSRTCDRLARVEEKLK 1130

```

Db 971 LNOETIKNOAKTKILCELOQTQDTQVTQANKQSQEVLTKTSLAHLKSKVCCELOKLE 1030
QY 1131 EKS--QOQOEKQOQLNVQSEMSQKINEIENLK-----NELAKNLTLEHMET-- 1179
Db 1031 KQSEDEKISLOSDIGFISECCLSMELKLADIWNQABELRPDQOQESGVELQHHSTTA 1090
QY 1180 -EKELEAOKLNNYEVKSIKERVKLKELQKSFETERDHLR-----GYI 1223
Db 1091 EESLNVKPTQEQOTERTLTTEYERRI--EQLEESLQRAQSELSILEKXKTDENKSLQLEYM 1149
QY 1224 RETEATGLQKEELKAHIIHLKEHOETIDELRSVSSEKTAQIINTQDLEKSHTKLOBEIP 1283
Db 1150 AKTISENENRSFRAYCLDKTKRYEQLOQTNEKLASVT-----TQCQVHLD 1200
QY 1284 VLHEEQELLPNVKVSQETQMTNELELLTQSTTKDSTTLARIEMERLR--LNEKQESQ 1341
Db 1201 VIKRSLQ-----EKITQAEKRNEL-----AVHKAELEKIRETLKEKSSYK 1243
QY 1342 EETKSITKTRDNLUKTIKEALEVKHDQKHEHIRETLAKIQESQSQKQOSLNMKEKONETTK 1401
Db 1244 EKLQAEERD--KEISR-LEV-----MRNTIAELHKTNSDREVEL-----EGVK 1285
QY 1402 IVSEMEQFKPK---DSALLRIETEMGLSKRLQESHDMSKSAKVK--DDLQRL----- 1450
Db 1286 ----MEKCOLKLYDKSML--ELEQ-LQCTAD--QKSSDLLPGSSNENIDDLQKCKDQYVQD 1338
QY 1451 QYVQESQDLKENIKEIVAKH-----LLETELKVAHCKLQEQETIN 1494
Db 1339 LELLRGEKAEELSEIQKINGOHSNTIKLEETAEEMITLTQKELB--RCEIAKLETPK 1396
QY 1495 EURNVLSKEETEISTQKOLEA-----INDKLQNKIQBIYEKERQ 1534
Db 1397 SKEADIKI---ALHCAQLRLHAYDKLVCEYERLKGCLSDSNKLSNQLNKKVERLHAQLA 1453
QY 1535 L-----NKKQI--SEVQENVNELKQFKEHKAQXSALQSIESTOMLSTNRLQF----- 1580
Db 1454 LQEGISGRSEITKQGRSELKDAIDENKTVREAKVGLNSLKAVQENNSAQEQFKQKIAD 1513
QY 1581 ---SBEIQIMIKEEMKRVQEQALQIERDOLKENTKEIVAKMKSQEKYQFLKMTAVN 1637
Db 1514 IKGSDVDELQIKLSQEV-----RDHLSENEELKRLKQAEQLQNMVMDKERKLN 1563
QY 1638 ETQEKMKETIEHLKEQF---ETQKINLENTETENIRLTOI--LHENLEEMSVTKERDILRS 1693
Db 1564 SS-----LRQDFDKLEQTKLDLE-----EQLRAKKVEIDRRSKELGEVTKDCENIRS 1610
QY 1694 -----VEETLKVERDOLK---ENLRETITRDLKQELKIVMHLKEHOETIDK 1739
Db 1611 DLEAQTNDFLKERETINUTISDLRLHNEQLLETSTKNYLSDDITAANNLNLNEMKNLHDLTK 1670
QY 1740 ----LRGIVSEKTNISNMQKLEHSDNALKQAQDLKIQELRIAHMHLKEQOETIDKLRG 1795
Db 1671 ECKSLRSDLSQKEEYFQTKQLLDDETINLKENRYMEEKLSSGNALKEDCE---KLRS 1727
QY 1796 IVSEKTKLSNMQKOLENSNAKLOBK-----IQELKANEHQILITLKK-----DVNE 1841
Db 1728 TLESKELILOQKQLEERLTVINEKNGKQALLDAQLKSNETAFSLAKAWIKQSALATEA 1787
QY 1842 TOKKVSMEQO-LKKQIKDOOSLTLKLEIENLNAQELHENLEEMKSVMKERDNLRR----- 1896
Db 1788 ANKRSLEMEQKVDKRTREYELRSLTKTREINFRSEKERMDGTISSLLEDKRNLEKLC 1847
QY 1897 VEETL-KLERDLKESLQETKARDLEIQOELKTAARMLSKEHKEETVDLREKISEKTIQIS 1955
Db 1848 VTELLAKLRELPALHTOKVNGDVSI--ELNSNG-----SPTPAAPV 1889
QY 1956 DIQKDLD--KSKDELQKTKQELQKLEQLLRVKEDVNMGSHKINEMQLKKQF 2006
Db 1890 ATKPKPLDCNSAECVPKSSSLETAE-----RKNRRMTAVDENRKQF 1930

```

RESULT 7

```

AAU83007
ID AAU83007 standard; protein; 240 AA.
XX AAU83007;
XX
DT 23-APR-2002 (first entry)
XX
DE Human homologue of YPD1 protein target for antifungal compound.
XX
KW Antifungal; fungal gene transcription; RPC34; POP3; TFA2; NAB2; MPT1;
MTR2; BOS1; POL30; RSA2; SQT1; MTW1; TPB1; SPC98; BFR2; RNAL1; GCD7; SKI6;
NIP1; LCP5; NCE103; ECO1; ORC2; CNS1; YPD1; TIM10; SRB4; yeast; fungus.
XX
OS Homo sapiens.
XX
PN WO200202055-A2.
XX
PD 10-JAN-2002.
XX
PF 28-JUN-2001; 2001WO-US020592.
XX
PR 29-JUN-2000; 2000US-0215164P.
PR 10-AUG-2000; 2000US-0224457P.
XX
PA (ANAD-) ANADYS PHARM INC.
XX
PI Moore J, Buurman ET, Desilva T, Harris S, Komarnitsky S;
PI Mendillo M, Moore D, McCoy M, Sanderson K, Haq T, Zhu S, Long F;
PI Davidov E, Thompson CM;
XX
DR WPI; 2002-147962/19.
DR N-PSDB; ABK32895.
XX
PT Screening candidate antifungal compound for interaction with essential
PT protein, modulation of essential protein activity, binding to essential
PT protein, by contacting protein with test compound and determining
PT effects.
XX
PS Claim 1; Fig 79; 522pp; English.
XX
CC The invention describes a method of screening a candidate antifungal
CC compound for interaction with essential proteins (EP) or for modulation
CC of EP activity e.g fungal gene transcription. The proteins tested in the
CC invention include RPC34, POP3, TFA2, NAB2, MPT1, MTR2, BOS1, POL30, RSA2,
CC SQT1, MTW1, TPB1, SPC98, BFR2, RNAL1, GCD7, SKI6, NIP1, LCP5, NCE103,
CC ECO1, ORC2, CNS1, YPD1, TIM10 and SRB4 from S. cerevisiae, C. albicans
CC and human homologues. The method involves contacting a culture with one
CC or more test compounds and determining the effects on the growth or
CC viability of the culture of cells which preferably comprises fungal cells
CC or yeast cells. Preferably the identified compounds interact with, or
CC modulate (preferably inhibit) activity of C. albicans EP. The inhibitor
CC compounds identified by the method are useful for preventing or
CC inhibiting fungal, particularly C. albicans growth in culture or in a
CC mammal. The antifungal agents interact with essential fungal elements
CC that can be used to treat fungal infection by preventing the growth and
CC preferentially killing the fungi, but does not inhibit the biological
CC activity of mammalian homologues. This amino acid sequence represents a
CC target protein used to test the antifungal compounds, described in the
CC method of the invention
XX
SQ Sequence 240 AA;

```

```

Query Match      8.9%; Score 1183; DB 5; Length 240;
Best Local Similarity 100.0%; Pred. No. 9.1e-44;
Matches 240; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1801 TDKLSNMQKOLENSNAKIQEIKELKANEHQILITLKKDVNETQKVSMEQIKQIKQDS 1860
Db 1 TDKLSNMQKOLENSNAKIQEIKELKANEHQILITLKKDVNETQKVSMEQIKQIKQDS 60
QY 1861 LTLSKLEIENLNAQELHENLEEMKSVMKERDNLRRVEETLKLRRDQKESLQETKARDL 1920
Db 61 LTLSKLEIENLNAQELHENLEEMKSVMKERDNLRRVEETLKLRRDQKESLQETKARDL 120

```


QY 1921 EIOBELKTARMLSKHEKTVDKLREKISEKTIQISDIQKDLDSKDELOKQKIOBLOKCEL 1980
 Db 121 EIOBELKTARMLSKHEKTVDKLREKISEKTIQISDIQKDLDSKDELOKQKIOBLOKCEL 180
 QY 1981 QLLRVKEDVNMHSHKKNEMQLKKQFPNLYCKCEMDNFQITKXKLHESLEIRIVAKERD 2040
 Db 181 QLLRVKEDVNMHSHKKNEMQLKKQFPNLYCKCEMDNFQITKXKLHESLEIRIVAKERD 240

RESULT 8
 ADE56037
 ID ADE56037 standard; protein; 3259 AA.
 XX ADE56037;
 AC ADE56037;
 DT 29-JAN-2004 (first entry)
 XX Human Protein CAA53052, SEQ ID NO 1876.
 XX Human; pain; neuronal tissue; gene therapy;
 KW spinal segmental nerve injury; chronic constriction injury; CCI;
 KW spared nerve injury; SNI; Chung.
 XX Homo sapiens.
 XX WO2003016475-A2.
 XX 27-FEB-2003.
 XX 14-AUG-2002; 2002WO-US025765.
 XX 14-AUG-2001; 2001US-0312147P.
 PR 01-NOV-2001; 2001US-0346382P.
 PR 26-NOV-2001; 2001US-0333347P.
 XX (GHEO) GEN HOSPITAL CORP.
 PA (FARB) BAYER AG.

XX Woolf C, D'urso D, Befort K, Costigan M;
 XX WPI; 2003-268312/26.
 DR GENBANK; CAA53052.
 XX New composition comprising two or more isolated polypeptides, useful for
 PT preparing a medicament for treating pain in an animal.
 XX Claim 1; Page; 1017pp; English.
 XX The invention discloses a composition comprising two or more isolated rat
 CC or human polynucleotides or a polynucleotide which represents a fragment,
 CC derivative or allelic variation of the nucleic acid sequence. Also
 CC claimed are a vector comprising the novel polynucleotide, a host cell
 CC comprising the vector, a method for identifying a nucleotide sequence
 CC which is differentially regulated in an animal subjected to pain and a
 CC kit to perform the method, an array, a method for identifying an agent
 CC that increases or decreases the expression of the polynucleotide sequence
 CC that is differentially expressed in neuronal tissue of a first animal
 CC subjected to pain, a method for identifying a compound which regulates
 CC the expression of a polynucleotide sequence which is differentially
 CC expressed in an animal subjected to pain, a method for identifying a
 CC compound that regulates the activity of one or more of the
 CC polynucleotides, a method for producing a pharmaceutical composition, a
 CC method for identifying a compound or small molecule that regulates the
 CC activity in an animal of one or more of the polypeptides given in the
 CC specification, a method for identifying a compound comprising the one or more
 CC pain and a pharmaceutical composition comprising the one or more
 CC polypeptides or their antibodies. The polynucleotide or the compound that
 CC modulates its activity is useful for preparing a medicament for treating
 CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
 CC therapy). The sequence presented is a human protein (shown in Table 2 of
 CC the specification) which is differentially expressed during pain. Note:

CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic form directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 3259 AA;
 SQ
 Query Match 8.5%; Score 1127.5; DB 7; Length 3259;
 Best Local Similarity 20.9%; Pred. No. 4e-40;
 Matches 643; Conservative 596; Mismatches 1011; Indels 831; Gaps 143;

QY 87 QTASGKTYTNGSEHGLVPIRAIHDFIKIKFPDPREFLRVSYMYIYETITDLCGT 146
 Db 447 ETASQTSFPDYNEGTQAVTEENIASQKRVLENEKGLLSSIE-----LEELKAEN 501
 QY 147 QKMKPLIIRBDVNRNVVADLTEVVTSEMAKWKTKGKSRHYGETKKNQSRSH-T 205
 Db 502 EKLSSQITLLEAQNRTGEADREVSEISIVDIANKRSSAEES---GDVLENTFSQKHKE 558
 QY 206 IFRMILESREKGPSCNCEGSKVYSHLNL-----VDLAGSRAAQTAAGVR 251
 Db 559 LSVILLEMKEAQE-----ETAFKLQLOGKRAEADHEVLDQKEMQEGEGIAPIK 610
 QY 252 LKEGCNINRSLFILGQVIKKLSD-GQVGGFINYRDSKLTIRLQNSLGNPKTRIICTTP 310
 Db 611 MK-----VF-----LEDTGQDFPLMPNEESSILPAVEKEQASTEHQSR---TSEE 651
 QY 311 VSPDETTLAQFASTA-----KYMKNTPYNEVSTDEALLKRYRKEIMDLK-----KOLE 360
 Db 652 ISLND--AGVELKSTKQDGKSLSAVPIQIGCHQDE--LERLKSQILLELNFHKAQIY 707
 QY 361 EVSLETRAQAMEKDOQLAQLEEE---KDLLQKVQNEKIENLRLMTSSSTLQOELKAK 416
 Db 708 EKNLDEKAK--EISNLNQLIEEFKKNADNNSAFTALSEERDQLLSQVKELSMVTELRAQ 765
 QY 417 RKRRTVTCGLKINKMNSYADQFNIPNTITTKTKLSINLRLREIDRESCVESDVFSNTL 476
 Db 766 VK-QLEWNLAERQRRLDYESTQAHNDNLLTEQIHSLSEIA-----KSKDVKIEVLQNEL 819
 QY 477 DTLSEIENPATKLLNQ-----ENIESEL-----NSLRADYDNLVDVDEQLRTEKE 522
 Db 820 DDV-QLPFSEQSTLIRSLQSLQNKQSEVLEGAERVRHISKVEELSALSQKLEIYTKM 878
 QY 523 EMELKLEKNDLDEF-EALERK-----TKK-----DOEMQLTHEISNLK---NL 562
 Db 879 D-QLLLEKRRDVTETLQTTIEEKDQVTEISFMTKMWQLNEEKESLSGVEIKTLKEQLNL 937
 QY 563 VKHREVYNQDL---ENELSSKVE-----LUREKEDQTKKQJE 596
 Db 938 LSRAEEAKKQVEEDNEVSSGLKQNYDEMSPAQISKEELQHEFDLLKKENQRKRLQA 997
 QY 597 YIDSOK--LENIKMDLSVLSIESIEDPKQMKQTLFPAETVALDADKESAPLRSENLELKEK 654
 Db 998 ALINRKEQLQVSR-LEELANLKDSEKKEIPLSETERGEVEEDK-----ENKEYSEK 1049
 QY 655 MKELATTVKOMENDIQLYSQLEAKKKMQO---VDLEKEL--QSAFNEITK-ITSLIDGKV 708
 Db 1050 ----CVTSKQBEIYLVKQITSEKEVELOHIRKLEEKLAEEQFQALVKQMNTLODKT 1105
 QY 709 PKDLLCNLELEGKITDLQKELNKEVEENE---ALUREVILLS-----E 748
 Db 1106 NQIDLLQAEISENQALIQKLIITSDASDGSVALVKETVISPPTGSGSEHMKPELEEK 1165
 QY 749 LKSLSEVRLRKETQDKSEELHIITSEKDKLFSEVHVESRVQGLLEIGTKDGLAAT 808
 Db 1166 ILALEKEKEQLQKQLE-----AUTSRK-AILKKAQEKE---RHURELKKQKDDYNRL 1215
 QY 809 QSNYKSTDOEFONFKTLHMDFFQKVMVLEENRMQETVNIKSKEAQKFDSSIGALKTEL 868
 Db 1216 QSOFDQSENENIG-----DQLRQLQIVRESIDGKLP--STDQOESCSSTPGLEEPL 1267
 QY 869 SYNTQELQKTRVQERL-----NEMEQLKEQLENRDSPLQTVREK 910
 Db 1268 -FKATE-QHHTQPVLESNLCPDWPVSHSEDAALQGGTSAQIKAKLKE-----IAEAK 1318

KW Golgin-245; neurodegenerative disease; neuroprotective; neurotropic;
 KW antiparkinsonian; anticonvulsant; gene therapy; human; variant.
 XX
 OS Homo sapiens.
 XX WO2003087403-A2.
 XX
 XX 23-OCT-2003.
 PD
 XX 16-APR-2003; 2003WO-EP003958.
 PF
 XX 16-APR-2002; 2002EP-00008553.
 PR 16-APR-2002; 2002US-037242P.
 XX
 XX (EVOT-) EVOTEC NEUROSCIENCES GMBH.
 PA
 XX Von Der Kammer H, Pohlner J;
 PI WPI; 2003-845345/78.
 XX N-PSDB; ACF58150.
 DR
 XX
 XX Diagnosing or prognosticating a neurodegenerative disease comprises
 PT comparing a level or activity of a transcription or translation product
 PT of golgin-245 gene in a test sample with a reference value of a known
 PT disease or health status.
 XX
 XX Claim 18; Fig 5; Opp; English.
 PS
 XX The invention relates to diagnosing or prognosticating a
 CC neurodegenerative disease in a subject, or determining whether a subject
 CC is at increased risk of developing the disease. The method involves
 CC determining a level and/or an activity of a transcription or translation
 CC product of a gene coding for golgin-245 and/or a fragment, derivative or
 CC variant in a sample from the subject and comparing the level and/or the
 CC activity to a reference value. The composition and methods are useful in
 CC diagnosing, preventing and/or treating neurodegenerative diseases, such
 CC as Alzheimer's disease, Parkinson's disease, Huntington's disease or
 CC amyotrophic lateral sclerosis. The recombinant, non-human animal is
 CC useful for screening, testing and validating compounds, agents and
 CC modulators in the development of diagnostics and therapeutics to treat
 CC neurodegenerative diseases, in particular Alzheimer's disease. Specific
 CC antibodies are used for detecting the pathological state of a cell in a
 CC sample from a subject. The present sequence represents a human golgin-245
 CC splice variant 1 polypeptide
 XX
 SQ Sequence 2228 AA;
 Query Match 8.2%; Score 1091; DB 7; Length 2228;
 Best Local Similarity 22.4%; Pred. No. 9.9e-39;
 Matches 554; Conservative 498; Mismatches 856; Indels 564; Gaps 111;
 344 LKRYRKEIMDLKKOLEEVSLETRAQAMEKDQLAQLLEKDLLQVQNEKIENLTRLMTV 403
 DB 1 MFKKLQKISEEQQLQALAPAQASNSSTP-----TMRSR 38
 404 SSSLTIQ-QLKAKRRRTWICLGKINKMNSNYADQFNIP-----TNITKTH 451
 DB 39 TSSTFEQLDEGTPNRESGDTQSPAQLQLRVPSVESLFRSPIKESLFRSSKESLVRTSS 98
 452 KLSINLLRIDSVCVS---ESDVSFTLTLSEIENWPATKLNOENI-----ESELN 501
 DB 99 RESLNRL-DLDSSTASFDPSPDMSRAEDLVGN-----SDSLNKQLQLRLRMERSLS 151
 502 SLRADYDNLVLDYEQLRTEKEEMELKLEKNDLDEPEALERKTKQDQEMQLIHEISLNKN 561
 DB 152 SYRGKSELVTAQMLQREKKLQGLILSQSD-----KSLRRIAREELQMDQAK- 203
 562 LVKRVYNQDLNELSSKVELLRKEDQIKLQEVID--SQKLEN--IKMDLSYSLESI 617
 DB 204 --KH-----LQEEFDASLE---EKDQYISVLQTVSLKRLRNGPMNVVLKPLPOL 251
 618 EDPKQM-----KQTLFDAETVALDAKRESAFLRSENLE-LKEKKMELATTYKOMENDIQLY 672

252 EPQAEVPTKEENPESDGEPPVEDQ-----TSVKLTLETQORVKROENILKCRKETTQSH 305
 QY 673 OSQ-----LEAKKKQVLDLEKELQSA-----FNEITKL-TSLDGVKPKDOLLNLE- 719
 DB 306 KEQCTLLTSEKALQEQLEKLEKIKOLHMAEKTKLITQLRDA---KNLILQLODQK 362
 QY 720 GKITDLQKELNKEVEERNEALREVI--LLSELKSLPSEVERLRKEIQDKSELHIITSEK 777
 DB 363 GMVI---AETKQKHETLEMKKEEIAQIORSRIKQMTTQGEELREQ----- 404
 QY 778 DKLFSEVVKESRVQGLLEIGTKDOLATTQSNYKSTDOEF-QNFKTLHMDPKYKVMV 836
 DB 405 -----KKSRAAFEELEKALSTAQTEEARRLKAEEMDEQIKTIKTSSEERISL 455
 QY 837 LEENERMNOELVNLKSAQKFDSSIGAL---KTSELSYKTOELQKREVEVERLINE-MEQ 891
 DB 456 QQELSRVQEVVDMKKSS--EQIAKLQKLHEKELARKEQL---TKKLTREREFQEQ 510
 QY 892 LKQELNRSPL---QTVREKTLITEKLOOTLEEVKTLTQEKDDLQLOESLQIERD- 946
 DB 511 MKVALEKSQSYLKIQKEQESLAELELE-LQKKAILESENKLRDLQAEATYRTR 568
 QY 947 --QLKSDIHTVNNIDTQEOLRNALSLLK-QHOETINTLAKSIEEVSRLHME----- 998
 DB 569 ILESSLKSLQENKQSKDLAVHLEAKKNHKEITVMVEKHKTELESUKHOODALWT 628
 QY 999 ENTGETKDFQOKWGDIDKQDLEAKNTQTLTADVKNDEIIEQQRKIFSLIQEKNE-LQQ 1057
 DB 629 EKLQVLQOQYTEMKUREKCEQ---KETL---LKDEIILFQAH---LEEMNEKTLE 677
 QY 1058 MLESVIAKEQKLTDLKENIEMTIENQBELRLLAGDEL-KKQOEIVA---QEKNH----- 1107
 DB 678 KLDVKQTELESLSSELSEVLKARHKLLEELSVLKQDTDKMQELEAKMQEKNNHQQQVD 737
 QY 1108 AIKKEGELS--RT-----CDRLAEVVEKUKKESQQLQEQQLLVNQEEMSEMQKINEL 1161
 DB 738 SIIEKEVSIQTEKALKDQINQLELLKERRDKHLKEHQ-----HVE 780
 QY 1162 NLKNELKNKELTLEHMETELELAQKL-NENVEVKSITTKERKVLKELQKSPETERDHLR 1220
 DB 781 NLEADIKSESEGLQ-QASAKLDVFOYSQATHQYKAAVEQLAQLOOKLLDLETERILLT 839
 QY 1221 GYIREITATGLQTEELKIAHILKEHQETIDELERSVSEKTAQIINTOD--LEKSHTKL 1278
 DB 840 KOVAEVEAQKDVCTVELDAHKIQVQDLMQOLEKQNSEMEQVKVSLTVVYESKLEDGKQ 899
 QY 1279 QEEIPVLHBEQELLPNVKVSETQETMNELELLTQ-STTKDSTTLARTEMERLRLNEKF 1337
 DB 900 EQTKQILVEKENN---ILQMEGQK--KEIITLQKLSAKEDSIHILNEEYETKFNQ-- 952
 QY 1338 QESQSEIKSLTKERDNLKTIKEALEVVKHDOLKEHIRETLAKTOESQSKQESLNMMKEDN 1397
 DB 953 EKXMEKVKQAKBMO--ETLKKKLDQEAUKKLELNTALEL---SQEKQFNK---1002
 QY 1398 ETTKTVSEMEQFKPK--DSALLRIETEMGLSKRLQESHDE-----MKSAVEKEDD 1446
 DB 1003 -----MLENAQANSAGISDAVSRLNETNQEQIESLTVHRRRLNDVISIWEKKLNQAAE 1057
 QY 1447 LQRLQEVLOSBDQKENTKEIVAKHLETFEBELKVAHCCLKE---QESTINELRVNLSE 1502
 DB 1058 LQEIHEIQLQKEQEVAELKQKILLFGCEKEEMNKETIWLKEGEGVQDFTTINLQELQK 1117
 QY 1503 KETEISTI---QKOLEAINDKLNQKIQEIVEKEEQLNIKQISEVQENVNELQKFEHRKA 1559
 DB 1118 KSAHVNSLAODETKLKAHLEKLEVDLUNK-----SLKENTFLOQLVELKMLAEEDKR 1169
 QY 1560 KDSALQSIKSKMELNLRNLOESQEEIQIMIKEK-----EMKRVQEBALQIBRDQ 1608
 DB 1170 KVS-----ELTSKLTDTDEEFQSLKSSHEKSNKSLKSLFKKLSEELAIQLDI 1219
 QY 1609 LKENTKEIV-AKKKE-----SQEKYQFL-----KWTAVNET-QEKMCIEIHELKQFE 1654
 DB 1220 CCKTTEALLAKTNELINISSKTNAILSRISHCOHRTTKKGAALLIKTCTVSELEAQLR 1279

QY 2213 F-----PSIKTEFOVLNKRKEMTOFLBEMINTRFDIEKKGKGIQKENDRICQVNNFFN 2266
 Db 1847 LTCQLEQKIK-ELDSCLVROKQVHRVEMELTSKY--EKQ-ALQQMDGR-----1893
 QY 2267 NRIIIMNESTEFERSATISKWEQDLKSLKKEKNEKLFKNYOTLTKTSLASGAQVNPPTQ 2326
 Db 1894 NKPTLEENTEESKSHLVQPKLLSNMWA--QHNDLEFK-----LAGAREKQKL 1942
 QY 2327 DNKNPHVTSRATQTTTEKIRELNSLHEAKESAMHKSIIKMKQKELBVTND-----IIA 2381
 Db 1943 GKEIVRLQKOLRLMRKHEHQEILEIL---KKEYDQEREKIKQEQEDLEKHNSTLKQLMR 1999
 QY 2382 KLOAKVHESNCKLEKT-KETIQVLQDKVALGAKPYKEIEDLQMKLVKIDLEKMKNAKEF 2440
 Db 2000 EFNTOIAQKQGELEMTKETINKAQVBEALLESQBETNQLLKIAEKDDDKRTAKRY 2059
 QY 2441 EKEISATKATVEYQKQVIRILRENLRSSQQAQDTSVISEHTDPOPSNKPITCGGGGIVQ 2500
 Db 2060 EEILDARE---EEMTAKVRDLQVLELQKQYQOKLEQEE---NPGNDNVT-----2104
 QY 2501 NTKALILKSHIRLEKEISKLOQNEQILQKQNELLSNN---OHLNSNEVKTWKERTLKR 2556
 Db 2105 -----IMELQVQ-----LAQKTTILSDSKLKEQEFREQIHNLDRUKKY 2143
 QY 2557 EAHQVQTCNSP 2568
 Db 2144 EKNVYAITVGTGTP 2155

RESULT 12

ABR61600
 ID ABR61600 standard; protein; 2230 AA.
 XX ABR61600;
 AC ABR61600;

XX 15-JAN-2004 (first entry)
 DT Human golgin-245 splice variant 2 polypeptide.
 DE Golgin-245; neurodegenerative disease; neuroprotective; nootropic;
 KW antiparkinsonian; anticonvulsant; gene therapy; human; variant.
 XX Homo sapiens.

OS WO2003087403-A2.
 PN 23-OCT-2003.
 PD 16-APR-2003; 2003WO-EP003958.
 PF 16-APR-2002; 2002EP-00008553.
 PR 16-APR-2002; 2002US-0372424P.
 XX (EVOT-) EVOTEC NEUROSCIENCES GMBH.

XX Von Der Kammer H, Pohlner J;
 PI WPI; 2003-845345/78.
 DR N-PSDB; ACF58151.

XX Diagnosing or prognosticating a neurodegenerative disease comprises
 comparing a level or activity of a transcription or translation product
 of golgin-245 gene in a test sample with a reference value of a known
 disease or health status.
 XX Claim 18; Fig 7; Opp; English.
 XX The invention relates to diagnosing or prognosticating a
 neurodegenerative disease in a subject, or determining whether a subject
 is at increased risk of developing the disease. The method involves
 determining a level and/or an activity of a transcription or translation
 product of a gene coding for golgin-245 and/or a fragment, derivative or

CC variant in a sample from the subject and comparing the level and/or the
 CC activity to a reference value. The composition and methods are useful in
 CC diagnosing, preventing and/or treating neurodegenerative diseases, such
 CC as Alzheimer's disease, Parkinson's disease, Huntington's disease or
 CC amyotrophic lateral sclerosis. The recombinant, non-human animal is
 CC useful for screening, testing and validating compounds, agents and
 CC modulators in the development of diagnostic and therapeutic to treat
 CC neurodegenerative diseases, in particular Alzheimer's disease. Specific
 CC antibodies are used for detecting the pathological state of a cell in a
 CC sample from a subject. The present sequence represents a human golgin-245
 CC splice variant 2 polypeptide

XX Sequence 2230 AA;

Query Match 8.2%; Score 1091; DB 7; Length 2230;
 Best Local Similarity 22.4%; Pred. No. 9.9e-39;
 Matches 554; Conservative 498; Mismatches 856; Indels 564; Gaps 111;

QY 344 LLKRYRKEIMDLAKQLEEVSLFETRAQAMKQDLAQLLEKDLQKQVNEIENLMLVT 403
 Db 1 MFKKLQKQISEEQOOLQALAPAQASSNSTP-----TMRSR 38
 QY 404 SSSLTILQ-QLKAKRKRVTWCLGKINKMKNVADQFNIP-----TNITKTH 451
 Db 39 TSSFTQLDEGTNRSGDTQSPAQLQURVPSVLSFRSPIKESLFRSSKESLVRTSS 98
 QY 452 KLSINLLREIDSVCS---ESDVFSNTLDTLSBIENWPATKLANQENI-----ESELN 501
 Db 99 RESLNRL-DLDSSTASFPDPPDMSEADLVGN-----SDSLNKQLQLRLRMERSLS 151
 QY 502 SLRADYDNLVLYEQRLTEKEMELKLNKNDLDEFALERKTKDQEMQLIHEISLNKN 561
 Db 152 SYRGKYSLVYAYQMLQREKKLQGLISQSD-----KSLRRIAEELQMDQQAQ---203
 QY 562 LVKREVNQDLENELSSKVELLREKEDQIKKLOEVID---SOKLEN---IKMDLSVLSST 617
 Db 204 --KH-----LQEFDALE---EKQYISVLQTVSVLLKQRLNGPMNVDLKPLQL 251
 QY 618 EDPKQV---KOTLFDATVALDRAKRESAFLRSENLE-LKEKMKELATYKQEMDIQY 672
 Db 252 EQAEVFTKEENPESDGEFVVDG-----TSVKLETQQRVQKQENLLKRCETIQSH 305
 QY 673 QSQ-----LEAKKQVQDLEKELQSA-----FNEITKL-TSLIDGKVPKOLLNLELE- 719
 Db 306 KEQCTLLTSKEALQEDRLQELKIKOLHMAEKTLLITQLRDA---KNLIEQLEQDK 362
 QY 720 GKITDLQKELNKEVEENEAAREEVI--LLSLKSLPSSEVERLKEIQKSELHIITSEK 777
 Db 363 GWVI---AETKQMHETLEMKKEEIAQRSRIKQMTTQGEELRQ-----404
 QY 778 DKLFSVVHKSRSVQGLLEIGKTKDLDLATQSNVKSSTQDEF-QNFKTLHMDFKQYKMW 836
 Db 405 -----KEKSERAAFEELKALSTAQKTEEARPKLKAEMDEQIKTIETSEERISL 455
 QY 837 LEENRMNOEIVNLKSAQKFDSSIGAL-----KTELSTYKTOELQEKTRVEQERLNE-MEQ 891
 Db 456 QQELSRVQGVVDVMKXSS---EQIAKLQKLHKEELARKEQL---TKLQTREREQEQ 510
 QY 892 LKQELNDRGFL---QTVREKTLITEKLOQTLLEEVKTLTOEKDDLKQLESQIERD- 946
 Db 511 MKVALEKSQSEYLIKISQEKQSSLALEELE--LQKKAILTSESNKRLDQQAERTYR 568
 QY 947 --OLKSDIHTVNMNIDTQBLRNALLESK-OHOETINTLKSITSEVSNLHME-----998
 Db 569 ILELESSLEKSLQENKQSKDLAVHLEAKNKKEITVMVEKHTELESKLKQDQALWT 628
 QY 999 ENTGETKDFEQKMGVIGDKQDLLEAKNTQTLTADVKDNEIEQQRKIFSLIQEKNE-LOQ 1057
 Db 629 EKLQVLKQYQYTEMKLRBKCEQ---KETL---LKDKEIIFQAH-----IEWNEKTL 677
 QY 1058 MLESVIAEKEQLKTDLKENIEMTIENQERLILGDEL-KKQOEIVA---QEKNH-----1107
 Db 678 KLDVKQTELESLSSELSEVLKARHKLHEELSVLKQDQTKMKQLEAKMDEQKNHQQQVD 737

QY	1108	AIKKGELS--RT-----CDRLAEVVEKLKESQQLQKQOQQLLNQVEEMSEWQKKINEIE	1161	2108	MDDHYECLNRLSLDL--EKEIEFHRIIMKKLKKVLS-----	YVTYKIBEQ	2149
Db	738	SIKEHEVSIQTERKALKDQINQINLELLKERDKHLKBHOA-----HVE	780	1672	--SH---LSELNTYKLOEREREVHLEELKLSVSSQSETLIVPRSAKVAAYTEQEEADS	1726	
QY	1162	NLKNELKNKELTLEHMETERLELAQKL--NENVEVKSITKERVAKELOKSFETPRDHLR	1220	2150	HECINKFEMDFIDEV-----EKQK-----ELLKIQ-----	2175	
Db	781	NLEADIKRGEGLQ--QASAKLDVQYQYQATHEQTKAYEQLAQQLKLDLTERILLT	839	1727	QGCYQKTYEIKSVLQRLNTEKELQVRGOEKBETVSSHFFEMCQYQERLUKHEAAK	1786	
QY	1221	GYIREIEATGLQKEELKIAHLKBHQETIDELRRSVSEKTAQIINTOD--LEKSHTKL	1278	2176	-----HIQDCDVPSREL-----	2212	
Db	840	KQVAEVAQKQVCTELDAHKIQVQDLMQQLKQSEMEQKVKSLTQVYESKLEDGNCQO	899	1787	QHEQSMIGHLQEELEENKKYSLIVAQHVKEGGKNNIKAKQNLNVFDDVQKTLQEKE	1846	
QY	1279	QEEPIVHFEQELLPNVQVSETOETWNELELEITEQ--SITKDSFTTLARIEMERLRNKF	1337	2213	F-----PSIKTEFQOVLNKRKEMTQFLEEMLNTRFDIEKLKNGIQKENDRICQVNNFFN	2265	
Db	900	EQTQKQILVEKENM---ILOWRGQK--KTEIITQKLSAKEDSIHILNEIYETKFNQ--	952	1847	LTCQILEQKIK--ELDCLVRQKEVHRVEMEELTSKY--EKQ--ALQQMDGR-----	1893	
QY	1338	QBSQEEIKSLTKERDNLTKTKEALEVQHDQKEHIRETLAKIQESQSKQPOSQSLMMKEKN	1397	2267	NRIIIMNESTFEERSATISKWEQDLSKLKKNKELFKNVQTLKTSILASGAQVNPITQ	2326	
Db	953	EKKMEKVQAKENQ--ETIKKLLDQAKLKELENTALEL---SOKEQQFNK--	1002	1894	NKPTELLEENTEKSKSHLVQPKLLSNMEA--QHNDLEPK-----LAGAREKQKL	1942	
QY	1398	ETTIVSEMEQFKPK--DSALLRIEIMGLSKRLQESHDB-----MKSVAKKQD	1446	2327	DNKNPHVTSRATQITTEKIRELENSLHSAKESAMHESKIIIMQKELEVTND-----IIA	2381	
Db	1003	-----MLEMAQANSAGISDAVSRLETNQKEQIESLTVHRRRLNDVTSIWEKKLNQAAE	1057	1943	GKEIVRLQKDLRMLKREHQOELLEIL--KKEYDQEREKIKQOEDELELKHNSLTKQLMR	1999	
QY	1447	LQRLQVLUQESDQLKENIKIIVAKHLETEELAVAHCLKE-----QEEETINELRVNLSE	1502	2382	KLQAKVHESNKCIEKT--KETIQVLQKVALGAPKPYKEETIEDLKMVLKVLDEKKNNAKEF	2440	
Db	1058	LQEIHEIQLOEKEQEVABLQKILLFGCEKEEMNKETIWLKBEGVKQDTYINELQEQLKQ	1117	2000	EFTVQLAKQEOBLEMTIKETINKAQEVAELLESQOETNQLLKKIAEKDDDLKRTAKRY	2059	
QY	1503	KETELSTI---QKOLEALNDKLONKIQIIEYKEBQLNIKQISVQENNELKQPKFHRKA	1559	2441	EKEISATKATVEYQKEVIRLLRENLRSSQAQDTSVISHPTDPOPSNKPITCGGSGGIQV	2500	
Db	1118	KSAAHNSLAQDETKLKAHLEKLEVDLKN-----SLKENTFLOEQVLKMLAEDDKR	1169	2060	EEILDARE--EEMTAKVRDLQTLQELQKYQOKLEQEE--NPGNDNVT-----	2104	
QY	1560	KDSALQSTESKMLELTNRLOSQEEIQIMIKE-----EMKQVQEAQLIBERDQ	1608	2501	NTKALILKSEHIRELKEISKLQKQNEQLIKQKELLNSN-----QHLSNVKTWKERTLKR	2556	
Db	1170	KVS-----ELTSKLTUTDEBFQSLKSSHEKSNKSLDKSLFQKGLSEELAIQDI	1219	2105	-----IMELQIQ-----LAQKTLISDSKLKQEFREQIHNLEDLKKY	2143	
QY	1609	LKENTKEIV-AKKKE-----SQRKEVQFL-----KMTAVNET--QEKWCBIHLKEQFE	1654	2557	EAHKQVTCENSP	2568	
Db	1220	CKKTEALLEAKTNELINISSKTNAILSRISHCHQRTYKVEALLIKTCTVSELEAQLR	1279	2144	EKNVYATVGTTP	2155	
QY	1655	--TQKLNLENIENTENIRITQILHENLEEMRSVTKERDDLRSVEETLKVFE--RDQLENLR	1710	RESULT	13		
Db	1280	QUTEQNTLINSFQ--QATHOLEEKENQIKSMKADIESLVTEKEALQKEGGNQQAASEK	1337	ABR61601	standard; protein; 2250 AA.		
QY	1711	ETITRDLEKQBELKIVHMLKHBQETIDKLRGIVSEKTEINSMQKDLHSNDALKAQDL	1770	AC	ABR61601;		
Db	1338	ESCITQLKKE-----LSNINAVTLMKBELKEKKEVSISSKQL-----TDL	1379	DT	(first entry)		
QY	1771	KQEEELRIAHMLKQEQETIDKLRGIVSEKTDKLSNMQKDLNSNAKLOEKIQELKANEH	1830	DE	Human golgin-245 splice variant 3 polypeptide.		
Db	1380	NVQLOQNSIS-----LSEKEAAISLRKQYDEKCELLDQVQDL---SF	1419	XX	Golgin-245; neurodegenerative disease; neuroprotective; nootropic;		
QY	1831	QLITLAKDVNETOKKVSEMBQL-----KQIKQOQ-----SLTSLKLEIENLNLQAE	1876	XX	antiparkinsonian; anticonvulsant; gene therapy; human; variant.		
Db	1420	KVDTLSKE-----KISALEQVDDWSNKFSEWKKKAQSRPTQHTQNTVXELQIQLELKSKE	1473	OS	Homo sapiens.		
QY	1877	LHENLEBMSVMKERDNLRVETILKLERDOLKESIQETKARDLEIQBELKTKARMLSKEH	1936	FH	Key	Location/Qualifiers	
Db	1474	AYEKDEQINLLKEELDQONKRFPCDGLGEMEDDK--SOMEKKSNELETLSQYARIM---1528		FT	Misc-difference	2244	"encoded by TTT"
QY	1937	KETVDKLRKISEKTIQI---SDIQKDLKSKD-----ELQKKIQELQK--KEQLLRVKED	1988	FT	Misc-difference	2245	"encoded by ACT"
Db	1529	-----ELEDHITQKTIIESINEVLKYNQKQDIEHKELVQKLQHFQELGEEDKDNRYKE-	1582	FT	Misc-difference	2246	"encoded by TCA"
QY	1989	VNMSHKKINEMEQLKQFERNYLKCBMDNFQTKLHESLEIRIVAKER--DELRRIKE	2047	FT	Misc-difference	2247	"encoded by CCT"
Db	1583	---AEEKILTLE-----NQVYSMKAELE---TKK--KELEHVNLSVSKSEELKALD	1627	FT	Misc-difference	2248	"encoded by CGC"
QY	2048	SLWNERQFIATLRMIARDQNHQVPEKRLISDQOHLWESLREKXCSRIKELLKRYSE	2107	FT	Misc-difference	2249	"encoded by AGT"
Db	1628	RLESESAKLAELKR-----KAEQKIAIKQ--LLSQMEEK---EEQYKKGTE	1671	FT	Misc-difference	2250	"encoded by GGT"
QY				PN			WC02003087403-A2.

QY 2540 --OHLSEVKTWKERTLKRAHKQVTCENSP 2568
Db 2147 KEQEFREQIHNLLEDRLLKYEKNVYATTVGTP 2177

RESULT 15

ADE56031
ID ADE56031 standard; protein; 3187 AA.

AC ADE56031;

XX 29-JAN-2004 (first entry)

DE Rat Protein BAA05026, SEQ ID NO 1870.

XX Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.

XX Rattus norvegicus.

OS WO2003016475-A2.

PN 27-FEB-2003.

XX 14-AUG-2002; 2002WO-US025765.

XX 14-AUG-2001; 2001US-0312147P.

PR 01-NOV-2001; 2001US-0346382P.

PR 26-NOV-2001; 2001US-0333347P.

XX (GHEO) GEN HOSPITAL CORP.

FA (FARB) BAYER AG.

XX Woolf C, D'urso D, Befort K, Costigan M;

XX WPI; 2003-268312/26.

XX GENBANK; BAA05026.

XX New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.

PS Claim 1; Page; 1017pp; English.

XX The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (SNI)), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a rat protein (shown in Table 2 of
CC the specification) which is differentially expressed during pain. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 3187 AA;

XX SQ

Query Match 8.2%; Score 1089; DB 7; Length 3187;
Best Local Similarity 20.9%; Pred. No. 1.8e-38;
Matches 634; Conservative 558; Mismatches 1012; Indels 828; Gaps 124;
QY 113 IFOKIKKFPDREFILRVSYMEIYNTIT-----DILCGTQMKKPIIREDVNRVYVA 165
Db 221 VMORKLEEHAEALLGRAQVVDLLQKELTSAEQNQLD---SQQLLLEAEHSLRNTMEA 277
QY 166 DLTEVVYVYSEMALKWITGK-----SRHVG-----ETQNRSS 201
Db 278 ERQESKILMEKVELEMAERKEELYQLQGLERAGQAQAELEMOYGTTLQORHETEMEEK-- 335
QY 202 RSHTIFRMILESEKGEPSNCEGSKVSHNLNVDLAGSERAQAOTGAAGVRLKEGCNINRS 261
Db 336 ---TACISLQKNEQELQSACD-ALKEENSKLIQ-EQEQQAQAKSAQALQLEDE----- 384
QY 262 LFLIGQVIKLSQGVGFNYRD-SKI/TRLQNSLGNPKTRIICTITPVSPFETLTAL 320
Db 385 ---LQKSKELISQ-----FVNKNLEKHETSSQTSL-----PDVYNEGVQAV 423
QY 321 QFASTAKYMKNTPVYNEVSTDEALLKRYRKEIMDLKKOLEBS-----LETRAQAMEKDX 375
Db 424 MEESVASLQKR---VLEENKKGALLSSLELEIRAEKELCSRITLLEAQNRAEAGD 480
QY 376 LA-----QLLEE--KOLLQKQNEKIEMLTLMVTSSSLT-----LQOEL 413
Db 481 MVCEVSTAGIALNRSDSSTESQGVLENTFSQKHKELSVLLVEMKEAQEELAFKLSQL 540
QY 414 KAKRKR-----RVTVCLGKINMKKNS-----NYADQF 440
Db 541 QGKPEGDYEVLDREKQVMESEGLPSVTARDVLC---APRDKNKSPAVEGEQAGMRDOH 597
QY 441 NI-----PTNIT-----TTHKL-----SINLLREI 461
Db 598 GTLEAGPLNDTGMELNSPQPDVDKSLSAHVCLHQGELERLKTQVLESLTAKET 657
QY 462 DESVCSDESDFNTDLTSLTEWNPATKLLNQ--ENIESELSNLRADYDNLV-----LD- 513
Db 658 HEKNLSEK--AKEISSITQL-----TKFKESAEEARSTLTAVCEERDQLLYRVKELDV 709
QY 514 YBQLATEKEEMELKLEKNDLDEFBALERKTKKQEMQLIHEISNLKLVKREYVNDL 573
Db 710 LGELRAQVRELETSLAE--AEKQRGDYESQRAQHNLITQIHSLSIETRAKSKDVKIETL 766
QY 574 ENELSKVELLEKEDDQIKLQFYDSQKLENIKMDLSYLSIESIEDPKMKQTLFPAETV 633
Db 767 QRELDGVQLQFSEQGTQIKSLQSLQTKB-----SEVLGAERMK----- 806
QY 634 ALDAKRESAFLRSENLKELKMKELATTYKQMENTI-QLYQSQLEAKKQXQVDLEKELQS 692
Db 807 -----DISKEMEELSQAQSKQELFAQMDQLLEKKQDVET-LQOTIQE 849
QY 693 AFNEITKLTSLIDGKVPDOLLNLEEGKITDQKELNKEV-----BENEALREVILLSE 748
Db 850 KQQVIT-----ELFSMTKQVQVQNEEKFSGLSGVEIKTLKEQLSLSR 891
QY 749 L-----KSLPSEVERLRKIQDKSEELHITSE 776
Db 892 AEGAKREQVEDSGAESSPKHGPHESSAEPVCKEALQOELEWLRKSEQRKRLQALIS 951
QY 777 KDKLFSEVWHKESRVQGLLEEIGTKDDLATTSQNTKSTQDQFQNFKTLHMDPEQKY--- 833
Db 952 RKELLOKVKSLLEELAKVREE--STKDSL--RESEKRELEEDSKN-----KDDPEKYGTS 1002
QY 834 -----KMWLEENRMNQEIYVNI-SKEAQKFDSSLGALKTSLSYKTOELOKTRV-- 882
Db 1003 EWRELEVSRLTITSEKEVELEGIIRDLKKAEEELQALVQRM---TQDLQNTKQKLDL 1059
QY 883 -QRLNEM-----BOLKEQLNDRSPLOTV 906
Db 1060 LQBEITENOATIQKFTITGTDAGDGSANVKTSSVSPRAGGEGHWPKELEGK---IVDL 1116

QY 907 EREKTLITEKLOTLBEVKUL---TQKODLKOIQLQESLQIERDQLKSDIHDVTNNIDTQ 963
 Db 1117 EKEKTOLOKLOALISRKAILKKAQKE--KHLKEELKEQDAYR-HLQBPFGQSKEN 1173
 QY 964 EQLRNALLESKQHETINTILSKISSEV---SRNLHME-----ENTGETK 1005
 Db 1174 ENIRAPILQAKESTDQQLPGTQOQEPHTGSEGLSLEGTEPASESDLHAAQPSHPGETA 1233
 QY 1006 D-----BFQOMVGID-KKQDLEAKNTOTLTADV--KONETIEOQRKFSLIQKNE 1054
 Db 1234 TLQATVSVAQIQDOLKEIEVEKEELEKISST-TSELTAKSEVILLQOEINEGLEION 1292
 QY 1055 LQOMLSVIAEKQLKTDLENMTTENIOFEURLLGDLELKQOEIQAQKNHAIKKEGE 1114
 Db 1293 LKAASHEAKAHTQLQOEL-ESSQLKIADLEHLKTLQPELETLQKHVGQKEEVSVYLVGQ 1351
 QY 1115 LSRTCDLRAEVEKLEKESQ-----OLQEQKQOQLLNQOESMSOKKINEIENL 1163
 Db 1352 LGEKEQTLTVQTEMEGEERLIKALHTQOLEMOKAHEERLKQVVEICELKQKPELE-- 1409
 QY 1164 KNEKKNELTLEHMETERLELAQKLENYEVEKSIITKERVKLKELOKSF-----E 1213
 Db 1410 -BESKAKQOLQKLAALISRKALKENKSLQELQSARDVAEHLTKSLADVESQVSVQN 1468
 QY 1214 TERDHLRG-----YIRIEATGLQTK-----ELKIAHILHKEHOETIDELRR 1256
 Db 1469 QEKDALLGLALLQEERDKLIVEMDKSLLENQSLGGSCESLKLALGLTEDKEKL--MKE 1526
 QY 1257 SVSEKTAQIINTQLEKSHTKLOEIPVL-----HEEQELLPN 1294
 Db 1527 LBSVRCSKIAESTEWQEKHELOQEYEVILLQSVYNSNEABRIQHVVESVROKQEVYAK 1586
 QY 1295 V-----KKVSETQFTMNELE-----1309
 Db 1587 LRSABSDKREKQOLQDAQOEMEEMKMKRPAKSKQKILELEENDRLRAPAQVGA 1646
 QY 1310 -----LLTEQSTTKDSTTLARIEMRLNKNKPOESQEBEISLTKERDNLKTIKEALRV 1363
 Db 1647 NESMEALLSSNALKEE--LERITILEYKTLSEFEALMAEKNTLSEETNLKLOVEAQEL 1704
 QY 1364 KH-----DOLKEHIRE-TLAKIOESQSQEQSLNMEKDNETTKIVSEMEQPKPD 1413
 Db 1705 QOASLETTEKSDPKDVIETVAVVGSQEQSDLSSENAKJEDAEATLLANSA--KP-- 1759
 QY 1414 SALLRIEIMLGLSKRLQESHDEMSVAKEDDLO-RLQEV-LQESD-----OLKENIKE 1467
 Db 1760 -----GVSETF-SSHDDINNYLQOLQKGRITAELEMEKQKORELSQTTLENNKN 1807
 QY 1468 IVAKHLETEE-ELKVAHCLKEQEBETINELRVNLSEKETEISTIOKQLEAINDKLQNKIQ 1526
 Db 1808 ALLTOISAKDSKL-----LEBEVAKINMLNQIQBELSRVTKLBTAEERKDDLEERL- 1862
 QY 1527 BIYEKEEQNLKQISEQENYNELQFKEHRKAKDSALQSTESKMLELTNRLQESQEBIQ 1586
 Db 1863 -----MNQIAELNGSIGNYYQDVTDAQKN---EQUESEMQLNRCVSELEEBEQ 1909
 QY 1587 IMIKE-----KEEMKRVQEA-----LOIERDOLKENTKEIIVAKMKESEKQEQYQLK 1632
 Db 1910 QLNVKEKTVESEIRKEYNEKIQGAKGPGSKIHAKELQELKKEQOQEVKQKQKOCIRYIG 1969
 QY 1633 MTAVNETQEKCE-----IHLKE-----QFETQKLNLENTENIR 1669
 Db 1970 RISALEKTVKALEFVHTESQKDLQATKGNLAQAVEHHKKAQAEELSSFKILLDDTQSEAR 2029
 QY 1670 LTOIHLHENLEEMRSVTKERDDLSVEETLKVERDOLKENLRETIITRDLEKQEBELKIVMHM 1729
 Db 2030 ---VLADNL-----KUKELQSNKESIK---SQIKQK-DEDLRLRLEQAE-----KH 2070
 QY 1730 LKEHQETIDKLRLGIVSEKTEINISNMOKDLEHSDNALKAQDLKIQBELRIAEHMLKEQOET 1789
 Db 2071 RKEKKNMQEKLDALHREKA-----HVEDTL-----AEIQVSLTRKDKMKELQOS 2115
 QY 1790 IDLGRGIVSEKTDKLSNMOKLEN--SNAK-----LOEKIQELKANEHQILTLKCO 1838

Db 2116 LDSTLAQIAAFTKSMSSLODDRDRVIDEAKWQORFGDAIQTKEEVRLKEENCTALK-- 2173
 QY 1839 VNQTKKVSMEQOLKQIKQDSLTLSKLEIENLNAQELHEN--LEEMKVMKERNLRR 1896
 Db 2174 -DQLRQMTIHMEELK-----ITVSRLEHD-----KEIWESKAQTELQHOQKAYDKLOE 2220
 QY 1897 VEETLKLRQOLKESLOETKARDLEIQOELKTARMLSKHEKHEVTDKLRE-----KI 1947
 Db 2221 ENKELASQLEEAQGLYHDSKNELTKLESELKSLKDQSDTLKNSLEKCREHNNLEGIKQ 2280
 QY 1948 SEKTIQ-----ISDIOKDLKSDK---ELOKKTQLOELKLOLLRVKE-----DVMN 1991
 Db 2281 QERADIOCKNCQOLETDLTASRELTTRRDEHINVKEQKTIISLSGKEEAIOVAIAELHQ 2340
 QY 1992 SH-KKINEMBOLKQKQEPNYLCKCEMDNFOLTKKLHESLEIRIVAKER-----DEL 2043
 Db 2341 QHSKEIKELENLALSQEEENL-TLEBENKRAVEKTLQLTALETIKKESLEQKALDSFV 2399
 QY 2044 RIKESLAKMERDQIATLRMIARDRONHVQPEKRLSD--GOHLMESLREKCSRIKEL 2101
 Db 2400 KSMSSLODDRDRIVSDYQOL-----BERHLSVILEKDELIQDAAENNNKKEE 2447
 QY 2102 LKRYSEMDDHYECLNRLSLDLEKE-IEFHRIMKKL-----KVLVSVTYKKEQOHE 2151
 Db 2448 IR---GLRGHMDLNSENAKLDAELIOYRRDLNEVITIKDSQORQLLEAQLOQNKELRNE 2504
 QY 2152 CINKFEMDFIDEVEKQKELLIKIQLHQDCDVPSELRDLKLNQNMMDLHIEELKDPSES 2211
 Db 2505 CV-KLEGRLLKSEABKQSLQMSLDLALQENQGLSKSEIKSF-K-EQLTALHEEGALAVY-HA 2561
 QY 2212 EFPSTIKTEQOVL-----SNRKEMTOFLEWLNLNTRFDIEKKNIGIKENDRIQOVNFFNN 2267
 Db 2562 QURVREVEEQKLTAAALSSQKRTVDIQLBELVCQKEASKVSEIEDKLR--ELKHLHN 2619
 QY 2268 RIIAIMNESTETEERSATISK---EWEQDLKSLKKEKNEKLFKNYQTLTKTSLASGAQVNPT 2324
 Db 2620 AGI-MRNETETAEERVAELARDLVEMEQKLTVTTKENKOLTAQIOAFGKSMSS----- 2671
 QY 2325 TQDNKNPHVTSRATQUTTEKIRELENSLHEAKESAMHKSKEKIKMKKELEVNDIIAKLQ 2384
 Db 2672 LQDSRD-HATELSDL-----KKYDASLKLALQKGRQDUGRESVDLSQA 2717
 QY 2385 AKVHESNKLEKTKETIQVLQDKVAKGAPYKEEIEDLMMKLVKIDLEKMKNAKEFEKE- 2443
 Db 2718 FPL-----TTSSENI-----SSRLEKLNQQLISKDEQLHLSSLESSH 2755
 QY 2444 ---ISATKATVEYQKEVIRLLR--ENLRSQQAQDTSVISEHTDPOPNKPLTCGGGSGI 2498
 Db 2756 NQVQSFTKAMTSLQNERDHLWNELEKFRKSEBKORSAA-----PSAASSPAE----- 2803
 QY 2499 VONTKALI--LKSEHIRLEKEISKLQOQNEOLIKQKNEL-----LSNQHLSNEVKTWK 2550
 Db 2804 VOSLKAMSSLQNDRRDLKELKNLQOQYLOMNOEITELRPLKQALQESQDQTKALQMK 2863
 QY 2551 ER-----TLKREAHKQVTCNSPKPKVTGTASKKQITPSQCKERNLODPV 2597
 Db 2864 FELQENLSWQHELDQLRVEKNKSWELHERRMKEQVYMAISDKDQQLGHLQNLREL----- 2919
 QY 2598 PKESPCKSCFFDSRSKSLPSPHPVYFDNSSLG 2629
 Db 2920 -----RSSOTQILPTQYQRASSG 2939

Search completed: July 29, 2004, 09:35:09
 Job time : 180.292 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 09:26:36 ; Search time 25.5784 Seconds
(without alignments)
5421.082 Million cell updates/sec

Title: US-10-045-631B-88

Perfect score: 13329

Sequence: 1 MAEGAVAVCVRRPLNSRE.....SQGPWHASSGKDVPCKTQ 2663

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	13329	100.0	2663	1	CENE_HUMAN	Q02224 homo sapien
2	1127.5	8.5	3259	1	G0B1_HUMAN	Q14789 homo sapien
3	1091	8.2	2230	1	G0A4_HUMAN	Q13439 homo sapien
4	1055	7.9	3911	1	A0A9_HUMAN	Q99996 h a-kinase
5	1038	7.8	3210	1	CENF_HUMAN	P49454 homo sapien
6	1032	7.7	2442	1	CEP2_HUMAN	Q9bvr73 homo sapien
7	959	7.2	2779	1	LVA_DROME	Q8m8s1 drosophila
8	956.5	7.2	2238	1	G0A4_MOUSE	Q91vw5 mus musculus
9	956.5	7.2	2867	1	RBP2_PLAVB	Q00799 plasmodium
10	956	7.2	1232	1	KF4A_HUMAN	O95239 homo sapien
11	952.5	7.1	1231	1	KF4A_MOUSE	P33174 mus musculus
12	935.5	7.0	1226	1	KF4A_XENLA	Q91784 xenopus lae
13	917.5	6.9	1225	1	KF4A_CHICK	Q90640 gallus gall
14	893.5	6.7	1957	1	SPOF_SCHPO	Q10411 schizosacch
15	894.5	6.7	1875	1	MLP1_YEAST	Q02455 saccharomyc
16	880.5	6.6	2869	1	RBP1_PLAVB	Q8wxh0 homo sapien
17	870	6.5	6885	1	SN22_HUMAN	P33176 homo sapien
18	867.5	6.5	963	1	KINH_HUMAN	Q07283 homo sapien
19	867	6.5	1898	1	TRHY_HUMAN	Q60282 homo sapien
20	866.5	6.5	957	1	KF5C_HUMAN	Q8mf91 homo sapien
21	863.5	6.5	8797	1	SNF1_HUMAN	P28738 mus musculus
22	863	6.5	956	1	KF5C_MOUSE	P12270 homo sapien
23	857.5	6.4	2349	1	TPR_HUMAN	P21613 loligo peal
24	856	6.4	967	1	KINH_LOLPE	Q15643 homo sapien
25	854.5	6.4	1979	1	TRIA_HUMAN	P47460 mycoplasma
26	854	6.4	1805	1	HMW2_MYCGE	Q61768 mus musculus
27	848.5	6.4	963	1	KINH_MOUSE	O94833 homo sapien
28	845.5	6.3	5171	1	BPEA_HUMAN	P17210 drosophila
29	844.5	6.3	975	1	KINH_DROME	Q81wj2 homo sapien
30	844.5	6.3	1583	1	GC02_HUMAN	P35978 strongyloce
31	844	6.3	1031	1	KINH_STRPU	P33175 mus musculus
32	837	6.3	1027	1	KINN_MOUSE	Q12840 homo sapien
33	835.5	6.3	1032	1	KINN_HUMAN	

34	833.5	6.3	1066	1	KL61_DROME	P46863 drosophila
35	833	6.2	3685	1	DMD_HUMAN	P11532 homo sapien
36	830.5	6.2	5938	1	MAC4_HUMAN	Q96pk2 homo sapien
37	829	6.2	1411	1	BEAL_HUMAN	Q15075 homo sapien
38	826.5	6.2	1955	1	PUMA_PARUN	O61308 parascaris
39	818	6.1	2871	1	DESP_HUMAN	P15924 homo sapien
40	816	6.1	1818	1	HMW2_MYCPN	P75471 mycoplasma
41	815.5	6.1	2022	1	ANTI_ONCVO	P21249 onchocerca
42	814.5	6.1	3678	1	DMD_MOUSE	P11531 mus musculus
43	812.5	6.1	8545	1	ANCI_CABEL	Q9n4m4 caenorhabdi
44	810	6.1	1038	1	KF17_MOUSE	Q99pw8 mus musculus
45	806	6.0	701	1	KF3A_MOUSE	P28741 mus musculus

ALIGNMENTS

RESULT 1
CENE_HUMAN STANDARD; PRT; 2663 AA.
AC Q02224;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Centromeric protein E (CENP-E protein).
GN CENPE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93024922; PubMed=1406971;
RA Yen T.J., Li G., Schaar B.T., Szilak I., Cleveland D.W.;
RT "CENP-E is a putative kinetochore motor that accumulates just before mitosis.";
RL Nature 359:536-539(1992).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=95196755; PubMed=7889940;
RA Thrower D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;
RT "Mitotic HeLa cells contain a CENP-E-associated minus end-directed microtubule motor.";
RL EMBO J. 14:918-926(1995).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=98437347; PubMed=9763420;
RA Chan G.K.T., Schaar B.T., Yen T.J.;
RT "Characterization of the kinetochore binding domain of CENP-E reveals interactions with the kinetochore proteins CENP-F and hBUBR1.";
RL J. Cell Biol. 143:49-63(1998).
RN [4]
RP FARNESYLATION.
RX MEDLINE=20459117; PubMed=10852915;
RA Ashar H.R., James L., Gray K., Carr D., Black S., Armstrong L., Bishop W.R., Kirschmeier P.;
RT "Farnesyl transferase inhibitors block the farnesylation of CENP-E and CENP-F and alter the association of CENP-E with the microtubules.";
RL J. Biol. Chem. 275:30451-30457(2000).
RN [5]
RP FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE KINETOCORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT AND/OR SPINDLE ELONGATION.
RN [6]
RP SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.
RN [7]
RP SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCORES DURING CONGRESSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.
RN [8]
RP SIMILARITY: Belongs to the kinesin-like protein family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; Z15005; CAA78727.1; --
DR PIR; S28261; S28261.
DR HSP; P17119; 3KAR.
DR Genew; HGNC:1856; CENPE.
DR GK; Q02224; --
DR MIM; 117143; --
DR GO; GO:0005699; C:kinetochore; TAS.
DR GO; GO:0005634; C:nucleus; TAS.
DR GO; GO:0008350; F:kinetochore motor activity; TAS.
DR GO; GO:0000067; P:DNA replication and chromosome cycle; TAS.
DR GO; GO:0007079; P:mitotic chromosome movement; TAS.
DR GO; GO:0007080; P:mitotic metaphase plate congression; TAS.
DR InterPro; IPR001752; kinesin_motor.
DR Pfam; PF00225; kinesin; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
DR PROSITE; PS00667; KINESIN MOTOR DOMAIN2; 1.
KW Motor protein; Cell division; Aff-binding; Coiled coil; Mitosis;
KW Cell cycle; Centromere; Lipoprotein; Prenylation.
FT DOMAIN 1 335 KINESIN-MOTOR.
FT 336 2471 COILED COIL (POTENTIAL).
FT 2472 2663 GLOBULAR (POTENTIAL).
FT NP_BIND 86 93 ATP (BY SIMILARITY).
FT LIPID 2660 2660 S-farnesyl cysteine.
FT SEQUENCE 2663 AA; 312087 MW; CEF13880C8C8C8 CRC64;

Query Match 100.0%; Score 13329; DB 1; Length 2663;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2663; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAEGAVAVRVPLNSREESLGETAQYVWTKDNNVYQVDGSGSFNDRVFGHNETTK 60
DB 1 MAEGAVAVRVPLNSREESLGETAQYVWTKDNNVYQVDGSGSFNDRVFGHNETTK 60
QY 61 NVYBIAPIIDSAIQNGTIFAYGQTASGKTYTMGSEDLGVIPRAIHDFOKIKFP 120
DB 61 NVYBIAPIIDSAIQNGTIFAYGQTASGKTYTMGSEDLGVIPRAIHDFOKIKFP 120
QY 121 PDRELLVSVMEIYNETITDLCGTQKMKPLIREDVNRNVYVADLTVEEVVYTSSEMALK 180
DB 121 PDRELLVSVMEIYNETITDLCGTQKMKPLIREDVNRNVYVADLTVEEVVYTSSEMALK 180
QY 181 WITGKSRHYGETKMNQSRSHITFRMILSRKSGPSPNCEGSKVSVSHLMLVDLAGSE 240
DB 181 WITGKSRHYGETKMNQSRSHITFRMILSRKSGPSPNCEGSKVSVSHLMLVDLAGSE 240
QY 241 RAAQTGAAGVRLKECINRSIFILGVTKLSDGVGGINVRYSKLTIRLQNSLGNP 300
DB 241 RAAQTGAAGVRLKECINRSIFILGVTKLSDGVGGINVRYSKLTIRLQNSLGNP 300
QY 301 KTRIICTTIPVSFDTLALQAPASTAKYMNTPYNEVSTDEALLKRYKEIMDLKKQLE 360
DB 301 KTRIICTTIPVSFDTLALQAPASTAKYMNTPYNEVSTDEALLKRYKEIMDLKKQLE 360
QY 361 EYSLTEQAAMEKQDLAQLEEKDLQKVNKEINLTMVTSLSLTQQELKAKRKR 420
DB 361 EYSLTEQAAMEKQDLAQLEEKDLQKVNKEINLTMVTSLSLTQQELKAKRKR 420
QY 421 VTWCLGKINMKNVADQFNIPNTITTKHLKSLNLRIDEVCSBDFVSNLTDLTS 480
DB 421 VTWCLGKINMKNVADQFNIPNTITTKHLKSLNLRIDEVCSBDFVSNLTDLTS 480
QY 481 ELEMNPATKLQENIESELNSIRADYNLDYEQLRTEKEMELKAKRNDLDFEAL 540
DB 481 ELEMNPATKLQENIESELNSIRADYNLDYEQLRTEKEMELKAKRNDLDFEAL 540
QY 541 ERKTKDQEMQLIHEISNLKLVKREVYNQDLENELSSKVELLREKEDQIKLQEVIDS 600

DB 541 ERKTKDQEMQLIHEISNLKLVKREVYNQDLENELSSKVELLREKEDQIKLQEVIDS 600
QY 601 QKLENIKMDLSYLSIESIEDPKMQKTLFDAETVALDAKRESAFLRSENLEKMKELAT 660
DB 601 QKLENIKMDLSYLSIESIEDPKMQKTLFDAETVALDAKRESAFLRSENLEKMKELAT 660
QY 661 TYQMENDIQLYQSOLBAKKMQVDLEKELOSANEITKLTSLIDGKVPKDLLCNLEEG 720
DB 661 TYQMENDIQLYQSOLBAKKMQVDLEKELOSANEITKLTSLIDGKVPKDLLCNLEEG 720
QY 721 KITDLOKELNKEVEENALREEVILLSELKSLPSEVERLKEIQDKSEELHIITSEKDL 780
DB 721 KITDLOKELNKEVEENALREEVILLSELKSLPSEVERLKEIQDKSEELHIITSEKDL 780
QY 781 FSEVVKESRVQGLLEIGTKDLDLATTQSNYKSTDOEFQNFKTLHMDFOKYVNLLEN 840
DB 781 FSEVVKESRVQGLLEIGTKDLDLATTQSNYKSTDOEFQNFKTLHMDFOKYVNLLEN 840
QY 841 ERMNOEIVNLSKEAQKFDSSLGALKTSLSYKTOLEKTRVQERLMEQLEKOLENRD 900
DB 841 ERMNOEIVNLSKEAQKFDSSLGALKTSLSYKTOLEKTRVQERLMEQLEKOLENRD 900
QY 901 SPLQTVREKTLITEKLOQTLLEEVKTLTQEKDLDLQKQESLQIERDQKSDIHDTVMNNI 960
DB 901 SPLQTVREKTLITEKLOQTLLEEVKTLTQEKDLDLQKQESLQIERDQKSDIHDTVMNNI 960
QY 961 DTQOLRNALLESKHOQETINTLKSITSEVSRNLHMEENTGETKDFQOMVGIDKKOD 1020
DB 961 DTQOLRNALLESKHOQETINTLKSITSEVSRNLHMEENTGETKDFQOMVGIDKKOD 1020
QY 1021 LEAKNTQTLTADVKNDEIIEQORKIFSLIQSKNELQOMLESVIAEKEKOTDLKENIEMT 1080
DB 1021 LEAKNTQTLTADVKNDEIIEQORKIFSLIQSKNELQOMLESVIAEKEKOTDLKENIEMT 1080
QY 1081 IENOEELRLDGLKQKQOEIVAQEKNHAIKKEGELSRCTDLAEVVEKLEKSKOOLQEQK 1140
DB 1081 IENOEELRLDGLKQKQOEIVAQEKNHAIKKEGELSRCTDLAEVVEKLEKSKOOLQEQK 1140
QY 1141 QOLLNVQEMSEMOKKINEIENLNKELNKELTLEHMETERLELAQKLNENYEVKSLTK 1200
DB 1141 QOLLNVQEMSEMOKKINEIENLNKELNKELTLEHMETERLELAQKLNENYEVKSLTK 1200
QY 1201 ERKVLKELQKSFETERDLRGVIRIEATGQTKKEELIAHHLKHEQETIDELRSVSE 1260
DB 1201 ERKVLKELQKSFETERDLRGVIRIEATGQTKKEELIAHHLKHEQETIDELRSVSE 1260
QY 1261 KTAQIINTODLEKSHTKLQEBIPVLHBEQELLPNVKVSETQETMNELELLTEQSTTKDS 1320
DB 1261 KTAQIINTODLEKSHTKLQEBIPVLHBEQELLPNVKVSETQETMNELELLTEQSTTKDS 1320
QY 1321 TTLARIEMERLRLNKEKFOESQEEIKSLTKERDNLKTIKEALEVHKDQKHEHRETAKIQ 1380
DB 1321 TTLARIEMERLRLNKEKFOESQEEIKSLTKERDNLKTIKEALEVHKDQKHEHRETAKIQ 1380
QY 1381 ESQSKOESLNKKEKNDNTTKIVSEMOFKPKDSALLRIEIMLGLSKRLQSHDEMKS 1440
DB 1381 ESQSKOESLNKKEKNDNTTKIVSEMOFKPKDSALLRIEIMLGLSKRLQSHDEMKS 1440
QY 1441 AKEXDDLOQLQVLOSQESDQKLENKEIVAKHLETESEELKVAHCCCKAQOETINELRVNL 1500
DB 1441 AKEXDDLOQLQVLOSQESDQKLENKEIVAKHLETESEELKVAHCCCKAQOETINELRVNL 1500
QY 1501 SEKETEISTIQOLAEINDKLQNKIQEYKEEEOQLNIQISVQENNELKOPKEHKKAK 1560
DB 1501 SEKETEISTIQOLAEINDKLQNKIQEYKEEEOQLNIQISVQENNELKOPKEHKKAK 1560
QY 1561 DSALQSIKSMLELTLNRLQESQEEIQIMKEKEMKRVQEQALQIERDQKLENKEIVAKM 1620
DB 1561 DSALQSIKSMLELTLNRLQESQEEIQIMKEKEMKRVQEQALQIERDQKLENKEIVAKM 1620
QY 1621 KESQSEKEYOFLKMTAVNETQEKVCEIHLKEFQEKMLNENIETENIRLTQLIHENLEE 1680

Db 1621 KESQKEYQFLKMTAVNTEQKMCIEIHLKEQFETQKLNLENITENIRLTQILHENLEE 1680
 QY 1681 MRSVTKERDDLSRVEETLKVERDQKENLRETTITRDLEKQEEELKIVHMLKEHQTIDKL 1740
 Db 1681 MRSVTKERDDLSRVEETLKVERDQKENLRETTITRDLEKQEEELKIVHMLKEHQTIDKL 1740
 QY 1741 RGVSEKTNESINMQKDLHSNDALKAQDLKIQELRIAHMLKEQOETIDKLRGIVSEK 1800
 Db 1741 RGVSEKTNESINMQKDLHSNDALKAQDLKIQELRIAHMLKEQOETIDKLRGIVSEK 1800
 QY 1801 TDKLSNMQKDLHSNAKLOKQIELKANEHQLITLKKDVNETQKQVSEMEQLKKIQKQDS 1860
 Db 1801 TDKLSNMQKDLHSNAKLOKQIELKANEHQLITLKKDVNETQKQVSEMEQLKKIQKQDS 1860
 QY 1861 LTLKLEIENLNAQELHENLEEMKSVMKERDNLRRVEETLKLERDQKESLOETKARDL 1920
 Db 1861 LTLKLEIENLNAQELHENLEEMKSVMKERDNLRRVEETLKLERDQKESLOETKARDL 1920
 QY 1921 EIQOELKTARMLSKHEKTVDKLREKISEKTIQISDIQKDLKSKDELQKKIQELOKQEL 1980
 Db 1921 EIQOELKTARMLSKHEKTVDKLREKISEKTIQISDIQKDLKSKDELQKKIQELOKQEL 1980
 QY 1981 QLLRVKEDVNMSHKKNEMEQKQEPNYLCKEMDNFQLTCKLHESLEEIRIVAKERD 2040
 Db 1981 QLLRVKEDVNMSHKKNEMEQKQEPNYLCKEMDNFQLTCKLHESLEEIRIVAKERD 2040
 QY 2041 ELRRIKESLKMEDQFIATLREMIARDRONHQVKEPKLLSDGOOHLMSLEKCSRIKE 2100
 Db 2041 ELRRIKESLKMEDQFIATLREMIARDRONHQVKEPKLLSDGOOHLMSLEKCSRIKE 2100
 QY 2101 LKRYSEMDDHYECLNRLSLDLKEIEFHRIKMLKXVLSYVTKIEEQHECINKEFMD 2160
 Db 2101 LKRYSEMDDHYECLNRLSLDLKEIEFHRIKMLKXVLSYVTKIEEQHECINKEFMD 2160
 QY 2161 IDEVEKQKELLIKIQHLQDCDVPSELRLDLKQNMMDLHIEILKDFSEFPSPKTF 2220
 Db 2161 IDEVEKQKELLIKIQHLQDCDVPSELRLDLKQNMMDLHIEILKDFSEFPSPKTF 2220
 QY 2221 QQVLSNRKEMTOFLEBWLNRFDIEKXNGIKOKENDRICOVNFFNRLIATMNESTEF 2280
 Db 2221 QQVLSNRKEMTOFLEBWLNRFDIEKXNGIKOKENDRICOVNFFNRLIATMNESTEF 2280
 QY 2281 ERSATISKWEQDLKSLKEKNEKLFNRYQTLKTSLSAGQVNPPTQDNKNPHVTSRATQ 2340
 Db 2281 ERSATISKWEQDLKSLKEKNEKLFNRYQTLKTSLSAGQVNPPTQDNKNPHVTSRATQ 2340
 QY 2341 TTEKIRELENSLHEAKESAMHESKIIKQKLEVTNDIIAKLOKVHESNKLKTKET 2400
 Db 2341 TTEKIRELENSLHEAKESAMHESKIIKQKLEVTNDIIAKLOKVHESNKLKTKET 2400
 QY 2401 IQVLQDKVALGAKPYKEETEDLKMVLKVIDLEKMKNAKEFEKISATKATVQKEVIRL 2460
 Db 2401 IQVLQDKVALGAKPYKEETEDLKMVLKVIDLEKMKNAKEFEKISATKATVQKEVIRL 2460
 QY 2461 LRENLRSSQAQDTSVISEHTDQPSNKLPTCGGSGIVQNTKALILKSEHRLKEISK 2520
 Db 2461 LRENLRSSQAQDTSVISEHTDQPSNKLPTCGGSGIVQNTKALILKSEHRLKEISK 2520
 QY 2521 LKQOQNEQLIKQNELLSNNQHLNNEVKTWKERTLKREAHQVTCNSPKSPKVTGTASK 2580
 Db 2521 LKQOQNEQLIKQNELLSNNQHLNNEVKTWKERTLKREAHQVTCNSPKSPKVTGTASK 2580
 QY 2581 KQITPSQCKERNLQDPVKESPKSCFFDSRSKSLPSPHPVRYFDNSSGLCPEVQAGAE 2640
 Db 2581 KQITPSQCKERNLQDPVKESPKSCFFDSRSKSLPSPHPVRYFDNSSGLCPEVQAGAE 2640
 QY 2641 SVDSQFGPWHASSGKDVPECKTQ 2663
 Db 2641 SVDSQFGPWHASSGKDVPECKTQ 2663

RESULT 2
 GOB1_HUMAN

ID AC Q14759; Q14398; STANDARD; PRT; 3259 AA.
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Golgi autoantigen, golgin subfamily B member 1 (Giantin) (Macroglolin)
 DE (Golgi complex-associated protein, 372-kDa) (GCP372).
 GN GOLGB1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=94187728; PubMed=7511208;
 RA Seelig H.P., Schranz P., Schroeter H., Wiemann C., Griffiths G.,
 RA Renz M.;
 RT "Molecular genetic analyses of a 376-kilodalton Golgi complex membrane
 RT protein (giantin).";
 RL Mol. Cell. Biol. 14:2564-2576(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=94257116; PubMed=8198703;
 RA Seelig H.P., Schranz P., Schroeter H., Wiemann C., Griffiths G.,
 RA Renz M.;
 RT "Macroglin -- a new 376 kD Golgi complex outer membrane protein as
 RT target of antibodies in patients with rheumatic diseases and HIV
 RT infections.";
 RL J. Autoimmun. 7:67-91(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=95100974; PubMed=7802676;
 RA Sonda M., Mismi Y., Fujiwara T., Nishioka M., Ikehara Y.;
 RT "Molecular cloning and sequence analysis of a human 372-kDa protein
 RT localized in the Golgi complex.";
 RL Biochem. Biophys. Res. Commun. 205:1399-1408(1994).
 CC -!- FUNCTION: May participate in forming intercisternal cross-bridges
 CC of the Golgi complex.
 CC -!- SUBUNIT: Homodimer; disulfide-linked.
 CC -!- SUBCELLULAR LOCATION: Golgi; membrane-associated.
 CC -!- DISEASE: Antigen in chronic rheumatoid arthritis and in the
 CC autoimmune disease Sjogren's syndrome.
 CC -!- SIMILARITY: Belongs to the golgin family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X75304; CAA53052.1; -;
 DR EMBL; D25542; BAA05025.1; -;
 DR PIR; A56539; A56539.
 DR PIR; I52300; I52300.
 DR Genew; HGNC:4429; GOLGB1.
 DR MIM; 602500; -;
 DR GO; GO:0000139; C:Golgi membrane; TAS.
 DR GO; GO:0005795; C:Golgi stack; TAS.
 DR GO; GO:0016021; C:integral to membrane; TAS.
 DR GO; GO:0007030; P:Golgi organization and biogenesis; TAS.
 DR GO; GO:0007030; P:Golgi organization and biogenesis; TAS.
 KW Golgi stack; Antigen; Coiled coil; Transmembrane.
 FT DOVAIN 1 3235
 FT TRANSMEM 3236 3256
 FT DOVAIN 3257 3259
 FT DOVAIN 48 593
 FT DOVAIN 677 1028
 FT DOVAIN 1062 1245
 FT DOVAIN 1301 1779
 FT DOVAIN 1828 3185
 FT DOVAIN 2420 2423
 FT DOVAIN 2993 2996
 FT DOVAIN 2993 2996
 FT POLY-SER.
 FT POLY-SER.


```

Db 2666 EKKQREKEKKRSPQDVEVLKTTTTELFHSNEESGFFNELEALRAB--SVATKAELASY-- 2720
QY 2022 TKKLHSLSEIRIVAKE-----RDLRIKESLK_MERDQFIATLREMIARD 2067
Db 2721 -KEAEKLOE-ELLVKETNTMTSLQDLSQVRDHLAEAKELSLKEDETEVOESKACM 2778
QY 2068 RQNHQVPEKRLSLSDGQHLMESLREKCSRIKELLKRYSEMDDHYECLNRLSLDLKKEIE 2127
Db 2779 FELPPIKLSIASQDTGTLKISSNQTPQI--LVKNAG-----IQINLQSECS 2825
QY 2128 FHRIMKKLKYVLVYVTKIEQHECINKFMDFIDEVEK-QKELLIKIHLQDQCDVPSR 2186
Db 2826 SEEVTEIISQTEKIEKQELHAAEILDMESRHSITETTLKREHYAVQVLLKECGTLKA 2885
QY 2187 ELRDLK-----NONDLHTEETLKDFSES---EFP 2214
Db 2886 VIQCLRSKVEGFYNNCFSTLCSGSGWGQGIYILTHSQGFDI-ASERGESSESATDSFP 2944
QY 2215 -----SIKTEFOQVLS-----NRKEMTQFLEELNTR-----FDIEKLKNGIQK 2253
Db 2945 KTKIGLLRAVHNEGQVLSLTSFSPYSGEDHSIQOVSEPWLEERKAYINTISSLKOLITK 3004
QY 2254 -ENDRICQVNN-----FFNRIITAINNESTE----- 2278
Db 3005 MQLQREAVDYDSSQSHESFSDWRGELLALQVFEERSVLLAAFRTELTAFTGTTDAVL 3064
QY 2279 ---PEERSATISKWEODLSLKEKNEK-LFNQVOTLTKTSLSAGVNPPTQDNKNP--- 2331
Db 3065 INCLEQRIQBGVEYQAEMLQKADRRSLSELQALHAQWNGKIIYLKREQSEKESQIE 3124
QY 2332 -----HVTSRAT-----QUTTEK---IRELENSLHEAK-----ES 2358
Db 3125 LLEYNIQKQSQMLEQVVELSSMKDRATELQELQSEKVMVAELKSELAOTKLETTLK 3184
QY 2359 AMHKEKIIKMKQ-----ELEVTNDIATKLQAKVHESNCKLEKKEKTIQVLDKVAL 2410
Db 3185 AQHKLKLEAFLEVKDTEVHLNDDTLASEQKSGRELQWALEKEKA-----KLGR 3237
QY 2411 GAKPKYKEIEDLKKVLKIDLEKMKNAKEFEKESATKATVEYQKEVIRLLRENLSQQ 2470
Db 3238 SEEDKELEDLKFSL---ESQKORNQ-----INLLLEQKQLINESQKIESQRM 3286
QY 2471 AQDTSVISEHTDPOPSKNPLFCGGSGIGVQNTKALIKSEHIREKEISKLKQONE----- 2526
Db 3287 LYDAQLSSE---QGRNLEL-----QVLLSEKVRIREMSSTLDRERLHAQ 3329
QY 2527 -----OLIKQKNELLNNQHLNVEKVTWKERTLK-----RE 2557
Db 3330 LQSSDGTGQSRPPLPSEDLKLEKQLEKHSRIVELLNETEKYKLDLSLQTRQOMEKDRQ 3389
QY 2558 AHKQVTCENSPKSPKVTGTASKKQITPSQCKERNLQDPVPKESPKSCFFDSRSKSLPSP 2617
Db 3390 VHRKYL-----QTEQANTEQKK-MHELSQKVEDLQRLQEKKQVYKLDLEGQRLQ-- 3441
QY 2618 HPVRYFDSNLSGLCEVQNAESVD-----SQCPWIASGK 2655
Db 3442 -----GIMQEFQKQELERBEKESRRILYQNLNEPTTWLSLTSR 3480

RESULT 5
ID -CENF HUMAN STANDARD; PRT: 3210 AA.
AC P49454; Q13171; Q13246;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE CENP-F kinetochore protein (Centromere protein F) (Mitosis) (AH antigen).
GN CENPF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast carcinoma;
RX MEDLINE=95348175; PubMed=7542657;
RA Liao H., Winkfein R.J., Mack G., Rattner J.B., Yen T.J.;
RT "CENP-F is a protein of the nuclear matrix that assembles onto
RL kinetochores at late G2 and is rapidly degraded after mitosis.";
RN J. Cell Biol. 130:507-518(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95379848; PubMed=7651420;
RA Zhu X., Mancini M.A., Chang K.-H., Liu C.-Y., Chen C.-F., Shan B.,
RN Jones D., Yang-Peng T.L., Lee W.-H.;
RT "Characterization of a novel 350-kilodalton nuclear phosphoprotein
RL that is specifically involved in mitotic-phase progression.";
RN Mol. Cell. Biol. 15:5017-5029(1995).
RN [3]
RP SEQUENCE OF 2194-3210 FROM N.A.
RX MEDLINE=95336446; PubMed=7612011;
RA Li Q., Ke Y., Kapp J.A., Fertig N., Medsger T.A. Jr., Joshi H.C.;
RT "A novel cell-cycle-dependent 350-kda nuclear protein: C-terminal
RL domain sufficient for nuclear localization.";
RN Biochem. Biophys. Res. Commun. 212:220-228(1995).
RN [4]
RP CHARACTERIZATION
RX MEDLINE=95370296; PubMed=7642639;
RA Zhu X., Chang K.-H., He D., Mancini M.A., Brinkley W.R., Lee W.-H.;
RT "The C terminus of mitotin is essential for its nuclear localization,
RL centromere/kinetochore targeting, and dimerization.";
RN J. Biol. Chem. 270:19545-19550(1995).
RN [5]
RP CHARACTERIZATION
RX MEDLINE=94437347; PubMed=9763420;
RA Chan G.K.T., Schaar B.T., Yen T.J.;
RT "Characterization of the kinetochore binding domain of CENP-E reveals
RL interactions with the kinetochore proteins CENP-F and hBUBR1.";
RN J. Cell Biol. 143:49-63(1998).
RN [6]
RP FARNESYLATION
RX MEDLINE=20459117; PubMed=10852915;
RA Ashar H.R., James L., Gray K., Carr D., Black S., Armstrong L.,
RN Bishop W.R., Kirschmeier P.;
RT "Farnesyl transferase inhibitors block the farnesylation of CENP-E
RL and CENP-F and alter the association of CENP-E with the
RN microtubules.";
RN J. Biol. Chem. 275:30451-30457(2000).
RN [7]
RP FUNCTION: Probably required for kinetochore function, involved in
CC chromosome segregation during mitosis. Interacts with
CC retinoblastoma protein (RB), CENP-E and BUBR1.
CC [8]
RP SUBUNIT: Homo- or heterodimer.
CC [9]
RP SUBCELLULAR LOCATION: Nuclear matrix (but not in the nucleolus),
CC reorganization to the kinetochore/centromere (coronal surface of
CC the outer plate) and the spindle during mitosis.
CC [10]
RP DEVELOPMENTAL STAGE: Gradually accumulates during the cell cycle.
CC [11]
RP PTM: Hyperphosphorylated during mitosis.
CC [12]
RP This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC [13]
RP EMBL; U19769; AAA82889.1; -
DR EMBL; U30872; AAA82935.1; -
DR EMBL; U25725; AAA86889.1; -
DR PIR; PC4035; PC4035
DR Genew; HGNC:1857; CENPF.
DR GK; P49454; -
DR MIM; 600236; -
DR GO; GO:0005699; C:kinetochore; TAS.
DR GO; GO:0005634; C:nucleus; TAS.

```


1786 VHQICDQAOQDNLNLDIEKITETGAVKPTGECGEGSPDNTNYPGEGDKTQGSSECSISEL 1845
 1302 -----QETWNELELLTEOSTTKOSTTL-----A 1324
 1846 SFGSPNALVPMDFLNGQEDIHNLQLVKETSNEENRLLHVEDRDRKVESLNNEMKELDS 1905
 1325 RIEMELRLNEKQF---ESQEEIKSLTKERDNLKTIKEALEVHKDQ-----1368
 1906 KUHLQEVQWTKTEACIELEKIVGELKKENDSLSEKLEVFSCDHQELLQVETSEGLNSD 1965
 1369 -----KEHIRETLAKIOESQSQOSQOSLNKMKENDNETTKIVSEMEQPKPKDQALLRI 1419
 1966 LEMHADKSSREDIGDVAKVNDSW--KERFLDV--ENELSRIRSEKASIE--HEALYLEA 2019
 1420 EIMELGLSKPLQSHDEMKS-----VAKKEDLOLQLEVLQSES---DOLKEN 1464
 2020 DLEVVOQTEKLCLEKDNENKQKIVCLLEELSUVTSERNQURGELDTMSKKTALDQLEK 2079
 1465 IKEIVAKHLTEEBE---LKVAHCCLKEQEBETINELRVNLSEKETEISTIOKQLEAIND 1519
 2080 MKE-KTQELSHQSECHLCIOVAEAEVKEKTELLQTLSSDVSELLKDKTHLQELQSLBK 2138
 1520 -----KLQNKIOEIIKEGEQNLNIKOISEVQENNELKQFKEHRKADS-ALOSIE 1568
 2139 DSQALSUTKCELENOIAQL-NKEKELLVKESESLOARLSE---SDYEKLNVSKALEAAL 2193
 1569 SKMELTNLROESOEETIOMIK-----EKEEMKRVQALOI-----ERDQKENTK 1614
 2194 VEKGEFALRLSSTOEEVHQLRGTEKURVTEADEKKQLHIAEKLKERENDSLKQVIE 2253
 1615 EIVAKMESQSEKYOFLKMTAVNETOEMKCMIEHLKEQFETQ-----KLNLNIEFTEN 1667
 2254 NLERELQSENGQ-----ELVILDAENSKAHEVETLKTQIEEMARSLKVFELDLVLRSEK 2308
 1668 IRLTQILHEN-----LEEMRSVTKERD---DLRSVET---LKVERDQLENLEET 1712
 2309 ENLTQIQEQQQLSELKLLSSPKSLLEEQAQBIQIKESKTAVEMLQNLKE-LNEA 2367
 1713 ITRDLKQEEELKIVMHL-----KEHQ--ETIDKLKG-IVSEKTEINSMQ--KDLHSN 1762
 2368 VAALCGDGEIMKATEQSDPPTEEBEHLRNSIEKURARLEADEKKQLCVLQKSEHHA 2427
 1763 DALKAADLKIOBELRIAHMHLKEQOETIDKLGRIVSEKTDKLSNNKQDLENSNAKLQEKI 1822
 2428 DLLKGRVENLRELEIA---RTNQE-----HAALBAENSKEGEVETLKAKIEGWT 2473
 1823 QELKANERHQLITLKKD---VNETQKK-----VSEMEQLKKQIKDOQSL 1861
 2474 QSLRGLGLDVVTIRSEKEDLTNELQKEQRISELEIINSFENILQKEQEKQVQMKES 2533
 1862 TLKSL---ELENL-LAQELHENLEEMKSVMKERNLRRVETTLERDQLKESIQETKA 1917
 2534 TAMEMLQTLKELNERNVAALHNDQEAQK---KEQNLSSQVECLEKAKQLLQGLDEAKN 2590
 1918 RDLQEQELKTAARMLSKEHKTVDKLRKIESEKTIQISDIQDKLQSKD-----1966
 2591 NVIVLQSSVNGLIQVEDGQKLEKDEISRLKNGIQOQEQVLVSKLSQVEGEHQLWKEQ 2650
 1967 -----ELQKKIQELQKELQLLURVEDVNMHKKI-NEMEQLKQFEPNLYCKEM 2016
 2651 NLELRNLTVLEQKIQVLQSKNASLODTTLEVLQSSYKNLENELELTK-----M 2698
 2017 DNFQTLKLHESLEETRIVAKERDELRLRIKESLKMERDQFIATLRMIARDRONHVKPE 2076
 2699 DKMSFVEKVN-----KMTAKETELQREHEMA-----QKTAELOE-----ELSGE 2738
 2077 KRLLSGQOQHLMSLEKCSRIKELKRYSEMDDHYECNLRLSLDLKEKEIEFRIMKLLK 2136
 2739 KNPRLAGLQLLLEIKSSQDKLQELTLENSLKKSLDCMHK--DQVEKE-----2785
 2137 YVLSYVTKIKEGHECINKFEMPFDIVBKQKELLIKIQLHQDCDVPS--RELRLDLKLNQ 2195
 2786 -----GKVRFE-----IAEQQLR-LHEAEKKHQAALLDNTNKQVEIQTYYE-----KLTS 2830

2196 NMDLHIEILKDFSESEFFSIKTEFOQLVSLNRKEMTQFLFEELWLNTRFDIEKLKNGIQKEN 2255
 2831 K-----ECLSS-QKLEIDLKSSKEELNNLSKATTOILEELKKTQMDNLKYVNLKKN 2884
 2256 DRICOVNFFNRITAINNESTEFERSATI SKNEFQDLKSLKKEKNEKLFKQVQTLKTS 2315
 2885 ERA-----QGMKLLIKSCQLEKEELQKELSQ-LQAAQEK-----2921
 2316 ASGAQVNPQDNKNPHVTSRATOLTTEKIRELENSLHEAKESA---MHKESKLIIMQKE 2372
 2922 ----QKTGTWMDTK-----VDELATTE-IKELKTELEKTEKDEADEVLDKYCSLLISHEK 2969
 2373 LEVNDIIAKLOAKVHESNKLCKETKETTQ--VLQDKVALGAKPYKEEIED-LKMKLVI 2429
 2970 LEX-----AKEMLETOVAHLCSQOSQSDSGSPGLGPVPGSPPIPSVTEKRLSSGQNK 3024
 2430 DLEKMKNAKEFEKEISATKATVE-YQKEVIRLRENLRSSQOQDTSVISEHTDQPSNK 2488
 3025 SGRQSSGIWENGPGTTPATPESFSKSKKAVMSGI---HPAEDT---EGTEFEPEGL 3077
 2489 PLTCGGSG-----IVQNTKALILKSEHIREK-EISKLQOQNEOLIKQKNELLS 2537
 3078 PEVVKGFADIPKTKTSPYILRRTMATRTPSLAAOKLALSPLSLKENLAESKKTAG 3137
 2538 NNQHLSEVTKWERTLKREAHQVTCENSGPKPKVTGTASKKKQITPSQCKERNLQDPV 2597
 3138 GSR--SOKVQ-----AQRSPVDSSTILREP--TTKSVPVNNLPERS 3175
 2598 PKESPSCFFDSRSKSLFSPHPVRYFNDNSLGLCPFVQNGAESVDQPGVHASSGKDV 2657
 3176 PTDSPREGLRVKGRGLVSP-----KAGLESKGSSE-----3205
 2658 PECKTQ 2663
 3206 -NCKVQ 3210

RESULT 6

CEP2 HUMAN
 ID CEP2 HUMAN STANDARD; PRT: 2442 AA.
 AC Q9BV73; O14812; Q60588; Q9H450;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Centrosomal protein 2 (Centrosomal Nek2-associated protein 1) (C-NAP1)
 DE (Centrosome protein 250) (Centrosome associated protein CEP250).
 GN CEP2 OR CNAP1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1), AND AUTOIMMUNE DISEASE.
 RC TISSUE=Cervical carcinoma;
 RX MEDLINE=98165428; PubMed=9506584;
 RA Mack G.J., Rees J., Sandblom O., Balczon R., Fritzler M.J.,
 RA Rattner J.B.;
 RA "Autoantibodies to a group of centrosomal proteins in human autoimmune
 RT sera reactive with the centrosome.";
 RL Arthritis Rheum. 41:551-558 (1998).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), INTERACTION WITH NEK2,
 RP AND SUBCELLULAR LOCATION DURING THE CELL CYCLE.
 RC TISSUE=Placenta;
 RX MEDLINE=98311641; PubMed=9647649;
 RA Fry A.M., Mayor T., Meraldi P., Stierhof Y.-D., Tanaka K., Niess E.A.;
 RT "C-Nap1, a novel centrosomal coiled-coil protein and candidate
 RT substrate of the cell cycle-regulated protein kinase Nek2.";
 RL J. Cell Biol. 141:1563-1574 (1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RP MEDLINE=21638749; PubMed=11780052;

RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baggeley C.L.,
 RA Bailey J., Barlow K.F., Bates K.N., Bead L.M., Beare D.M.,
 RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
 RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P., Clee C.M.,
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
 RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
 RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
 RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
 RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
 RA Leharlaio M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
 RA Marsh V.L., Martin S.L., McConachie L.J., McElroy K., McMurray A.A.,
 RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
 RA Phillimore B.J.C.T., Prachalingam S.R., Plumb R.W., Ramsay H.,
 RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,
 RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
 RA Swann R.N., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
 RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M., Thorpe A.,
 RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
 RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
 RA Rogers J.,
 RA "The DNA sequence and comparative analysis of human chromosome 20.";
 RL Nature 414:865-871(2001).
 RP SEQUENCE FROM N.A. (ISOFORM 3).
 RC TISSUE=Placenta;
 RX MEDLINE=2238257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Carninci P., Prange C.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Abramson R.D., Mullahy S.J.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
 RA "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences.";
 RT proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RL [5]

RP PHOSPHORYLATION, AND INTERACTION WITH NEK2 AND PPLICA.
 RX MEDLINE=21060765; PubMed=10880350;
 RA Helps N.R., Luo X., Barker H.M., Cohen P.T.W.;
 RT "NIMA-related kinase 2 (Nek2), a cell-cycle-regulated protein kinase
 localized to centrosomes, is complexed to protein phosphatase 1.";
 RL Biochem. J. 349:509-518(2000).
 [6]
 RP PHOSPHORYLATION DURING CELL CYCLE.
 RX MEDLINE=22135747; PubMed=12140259;
 RA Mayor T., Hacker U., Stierhof Y.-D., Nigg E.A.;
 RT "The mechanism regulating the dissociation of the centrosomal protein
 C-Nap1 from mitotic spindle poles.";
 RL J. Cell Sci. 115:3275-3284(2002).
 CC -!- FUNCTION: Probably plays an important role in centrosome cohesion
 during interphase.
 CC -!- SUBUNIT: Monomer and homodimer (Probable). Forms a complex in
 vitro with both Nek2 kinase and the PPLIC catalytic subunit of
 protein phosphatase 1 (PPL).
 CC -!- SUBCELLULAR LOCATION: Component of the core centrosome. In
 interphase cells, it specifically associates with the proximal
 ends of both mother and daughter centrioles. Associates with the

CC centrosome in interphase cells. In mitotic cells, it dissociates
 CC from the mitotic spindle poles. At the end of cell division, it
 CC reaccumulates at centrosomes.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=3;
 CC Name=1;
 CC IsoId=Q9BV73-1; Sequence=displayed;
 CC Name=2;
 CC IsoId=Q9BV73-2; Sequence=VSP_007372;
 CC Name=3;
 CC IsoId=Q9BV73-3; Sequence=VSP_007370, VSP_007371;
 CC Note=No experimental confirmation available;
 CC -!- TISSUE SPECIFICITY: Ubiquitously and weakly expressed.
 CC -!- PTM: Differentially phosphorylated during cell cycle.
 CC Phosphorylation may regulate association/dissociation from
 CC centrosome. During M phase of mitosis, C-terminal part is
 CC phosphorylated by NEK2, suggesting that it may trigger the
 CC dissociation from the mitotic centrosome. It is dephosphorylated
 CC in vitro by the PPL phosphatase.
 CC -!- DISEASE: Antibodies against CEP2 are present in sera from patients
 CC with autoimmune diseases that developed autoantibodies against
 CC centrosomal proteins.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; AF022655; AAC06349.1; --
 CC EMBL; AF049105; AAC07988.1; --
 CC EMBL; AL121586; CAB89415.1; --
 CC EMBL; BC001433; AAH01433.1; --
 CC Genew; HGNC:1859; CEP2.

CC Cell cycle; Coiled coil; Phosphorylation; Alternative splicing;
 CC Polymorphism.

DOMAIN	95	158	COILED COIL (POTENTIAL)
DOMAIN	244	352	COILED COIL (POTENTIAL)
DOMAIN	395	1172	COILED COIL (POTENTIAL)
DOMAIN	1243	2227	COILED COIL (POTENTIAL)
DOMAIN	2262	2376	COILED COIL (POTENTIAL)
DOMAIN	246	250	POLY-LIG.
DOMAIN	464	2171	GLN/GLU-RICH.
VARSPLIC	313	313	V -> F (in isoform 3).
VARSPLIC	314	2442	/FTid=VSP_007370.
VARSPLIC	863	918	Missing (in isoform 3).
VARSPLIC	995	995	Missing (in isoform 2).
VARIANT	995	995	/FTid=VSP_007372.
CONFLICT	120	120	Q -> H (in dbSNP:2296403).
CONFLICT	136	136	/FTid=VAR_015649.
CONFLICT	136	136	L -> I (IN REF. 1).
CONFLICT	365	365	E -> A (IN REF. 1).
CONFLICT	372	372	H -> L (IN REF. 1).
CONFLICT	509	509	D -> E (IN REF. 1).
CONFLICT	552	552	E -> D (IN REF. 2).
CONFLICT	757	757	S -> I (IN REF. 1).
CONFLICT	784	787	E -> A (IN REF. 1).
CONFLICT	1153	1153	EVTK -> DEPO (IN REF. 1).
CONFLICT	1246	1246	Q -> H (IN REF. 1).
CONFLICT	1513	1513	H -> L (IN REF. 1).
CONFLICT	2082	2082	L -> P (IN REF. 1).
CONFLICT	2345	2345	D -> N (IN REF. 1).
SEQUENCE	2442	281135	AA; 281135 MW; BC2B8A3607B8272 CRC64;

Query Match 7.7%; Score 1032; DB 1; Length 2442;
 Best Local Similarity 20.7%; Pred. No. 1.1e-20;
 Matches 562; Conservative 539; Mismatches 944; Indels 674; Gaps 112;

FT CONFLICT 2160 2160 E -> V (IN REF. 2).
 FT CONFLICT 2200 2200 L -> P (IN REF. 2).
 FT CONFLICT 2217 2217 A -> T (IN REF. 2).
 FT CONFLICT 2271 2271 E -> D (IN REF. 2).
 SQ SEQUENCE 2779 AA; 315897 MW; 1CB3965102018AEE CRC64;

Query Match 7.2%; Score 959; DB 1; Length 2779;
 Best Local Similarity 20.2%; Pred. No. 1e-18;
 Matches 569; Conservative 528; Mismatches 951; Indels 768; Gaps 117;

QY 152 LIREDVNRVYVADLTVEEYVYVSEMALKWITKXSRHYGETYKMQRSRS-----HT 205
 Db 305 LVLRERLAELVNDLLETT--RCELQEELTTAREQRORNLRLQEQOEKASRSPQSEAAHT 362
 QY 206 IFRMILE-SREKEPNSCEGSKVSHNLVLDL-AGSRAAQTGAAGVRLKEGCVNRSFLF 263
 Db 363 DAQVSAELAKQLQELTN-----QLADLQATNEELRQQAQAQKLVQTBDE-----406
 QY 264 ILGOVTKLSGQGVGFYINRDSKLTIRLQNSLGGNPKTRI-ICTIIPVDFDETLTALQF 322
 Db 407 IVSORLEELATTAQELLEQEQKSAAQNEELAEKTTTELNVNLNRLLEEKI-----461
 QY 323 ASTAKYMNTP-YNEVSTDEALRYRKEIMDKKOLEE-----VSLTEPAQAMEK---373
 Db 462 ---AQSRKPLFLEDHSDSAAKQMOEDIQQLKLDETNKANIKLKCKQAEKKLQ 518
 QY 374 -----DQLAQLLEKLLQK-----VQNEK-----IENL-----397
 Db 519 KFQSDGQQQASLADNEELQORIAVLEDEKQWQOLANMOEDDRQEQSTESNPLQLE 578
 QY 398 TRMLVTSSTLQLOELKAKR-----KRVVWCLGKINKMKNVYADQFNI---442
 Db 579 TIRLEQKLEQALQALLSSSSSAESIEIVERHLECLQRRPASGDAQEQKQVHP 638
 QY 443 -PTNIT--TKTHKLSINLLRIDESVCSDFVNTLDTLSEIENWPATKLLNENIESE 499
 Db 639 GPSHVSELTQEQ-----TEEDSSGETLSQLE-----RLEFTQERGE---678
 QY 500 LNSURADYDNLVD-YEQLRTEKEEMELKLEKNDLDEFEALERTKKQOEMLIHEISN 558
 Db 679 -----VLDKLEQSAENLQOLARLESS---SSLQLLQREKX-----LISSTST 721
 QY 559 LKNLVKREHVYNQDLE-----NELSKVYELLREKEDQTKKLE 596
 Db 722 SSNLQELSSMRSESVATLDAGEGPVLFKECKESLSKLNSELEYKANDROAKFN 781
 QY 597 YIDSQK-----ENIKMDLSYLESIEDPKMQKQTLFDAET-VALDAKRESAF-----LRS 646
 Db 782 ---SKLAKEAKNCHTQLSLELLHKVKEASTAVETVTVVAVTAPNGKALAEYEQUNA 838
 QY 647 ENLELK-----EKMKELATTY-----KOMENDIQLYQSOLEAKKQKQVLD 686
 Db 839 QNAELKAVISRRLQELDELRESYPTEAPLAVGSDSQREDEILQLOSQLEDAKSLQAEQ 898
 QY 687 EKELQSAFNEITKLSLIDGKVPKOLLCNLEEGKITDLO-----KELN--KE 732
 Db 899 ROQIEQVQVQIKELRQ-----TEAQOLQVARQSAEITQLOLQSEQPDOLLNKENSHKQ 954
 QY 733 VEENALREEVILLBLKSPSEVRLKEIQDKSEELHIITSEKDLSEVVHKESRVQ 792
 Db 955 LEOQTRIRRE--LEARAESLEGELSITQTVAEQKQOLTESSEHALNKLMLQESQAQ 1012
 QY 793 GLLLEELGKTD-----DLATTQSNVKSFTDQFQNFKTL-HMDFEQ 831
 Db 1013 EELREIRAKEDDPOLREARVSKSLVAQVRELTSSQETVDALNQIQEYQGLEHAHKE 1072
 QY 832 KYQWVLEENRMNQEIIVNLSKEAQFSDSISGALKTELSYKTOBELQKTRVQBRLEMEQ 891
 Db 1073 QFNK-RELREKLKYALNKKRTQ-----DNADLEQKQVBLTSQLOEQOE 1116
 QY 892 LKQLENRDSPLQTVREKTLITEKLOQTLQEEVKTLTQE-----KDDLKQLOES 940
 Db 1117 LVQKKEE-----VEREPIVDNHRVQLOQVQSKLNEDLKAKIHLNLRNDRALRQKQ 1169

QY 941 LQIERDQ-----KSDIHDVTVMNIDTQEOALNALESILKQOETINTLKSKEI SEEVSRLNH 996
 Db 1170 IQ-EQELIQERDAELODA---NLVSKL-LRRERQADQEVQLQOENSLREESIKLQ 1224
 QY 997 MEENTGETKDEFOQKMGVGDKKQDLKAKNTQTLTADVKNNEIIEBQKQKIPSLIQEKNELO 1056
 Db 1225 EHNLQORVNE--EPTAVEDLRQLEAKSKF-----EKSKELIKLRNATIOSIQ--RELQ 1276
 QY 1057 QMLESVIAPKEQKQKTDLKENIEMTIENQBELRLLGDELKQOEVIAQEKNNHAIKKEGE--1114
 Db 1277 QLOQDQDSEVHVNRNARAHEQLRLEKDAEITAL-----ROBILKLSRRA-AGEGDDT 1329
 QY 1115 LSTCDRLAEBEKLKESQOQKQOQLNVQEMSEMOKKINEIENLNKLNKELTLL 1174
 Db 1330 ITKTSQLESQOQQAESLQVARELQQLRVOLTAQEQHAL-LAQQVYASDRANPEWTI 1388
 QY 1175 EHMETERLELAQKLNENYEEVKSITKERKVLKELQSFETERDHLRGYIRIEATGLQTK 1234
 Db 1389 ARLETHEGIAKQLEDASVIESL-----EAQNTLEQAR 1422
 QY 1235 BELKIAHILKHEQETIDELRRSVSEKTAQIINTQDLEKSHSTKLOEIPVLHBEQLLPN 1294
 Db 1423 SA-----ALEBQAASQANQAAQKQVQILEQKQEBEQEQR-----QQOQLOER 1471
 QY 1295 VKKVSQETWN-ELELLTBSQSTTKDSTTLARIEMERLRLNEKFPQESQEEIKSLTKBRDN 1353
 Db 1472 FYELGREQAQSRLQELTSEA-----EESRQQLAGLRTYES 1509
 QY 1354 L-----KTKIEALEVKHDPQKHEHRETLAKTIOESKQESQOSLNMEK--DNETTKIVSEME 1407
 Db 1510 LLAHXSQLTATAQAREQMSHSSQELAEQLQDLVKEADLHRQVQYDAKAAKATELD 1569
 QY 1408 QFK-PKDSALIRIEMI GLSKRLOESHEDEKMSVAKKDDLQRLQEVLOESDOLKENIK 1466
 Db 1570 ELECDLNSHVRAAAETRELQOQLERSQE---LVAQTEELQRLNEEBQFVERESTLSR 1626
 QY 1467 EIVAKHLETEBELVAHCCLEKEQETINELRVNLSEKETEISTIQKLEAI-----ND 1519
 Db 1627 EVTL-----LRLQHDSEADQVLELQELRMAQMDKTEMNDLRTQIDALCANHSQELQ 1678
 QY 1520 KLQNKIQEI-----YEKEQNLTKQISEVQENNVNELKQFKEHRYAKKASQALS 1566
 Db 1679 ALQRIAEALDTGQNTDQVVIETENKRLAQLSELQALQARQHQOQOQHHPAVOS 1738
 QY 1567 IE-----SKMELTNRLQESQEIQIMIKEEMKRVQEAQLQIERTDQKENTKEI 1616
 Db 1739 QQHPPPSILFFGGDALAAPSPFDETAQPLRVSSLAASAPPPISPPPTIE--DLQNVSDL 1796
 QY 1617 VAKKESQEKVQFLKMTAVNETQKMCIEIHLKEQFETQKLNLENIETENIRLTQILHE 1676
 Db 1797 ---EKHAQDLE---TKLLARN-----QNLAEQ-EERRIQLEQLSEVERLLSERTQ 1840
 QY 1677 NLEEMRSVTKERDDLRSVEETLKVRE-----DQKENDRETITRLE-----1718
 Db 1841 QLADIQTANEERDLAALAEKLIQAPAAAPLDMFFGQAEETVPDASHHLDLGLPQTEPV 1900
 QY 1719 -----KQEBELKIVHMLKEH-QETIDKLRGIVSEKTNIEISN--WOKDLEHSDALKAQ 1768
 Db 1901 VEPLIQPKAYLCQPKQEIQEQTAQID--WGVDEDPWASAANEAPQDVEHLH-----1952
 QY 1769 DLKIQEBELIAHMLK---EQETIDKLRGIVSEKTDKLSNMQKOLENSAKLQKIQE 1824
 Db 1953 -----TRIAQLEQLSNABQKT-----ELQTKAAKMLKRLKE 1985
 QY 1825 LK-----ANEBHOLITLKKDVNETQKVSMBQKQIKQDQSLTSLKLE-----IENLNL 1873
 Db 1986 YKTKATTATPTVTDNDLST-----IIEELKHQLOQLESRLSKAEIISQOHALEKEL 2040
 QY 1874 AQEL-----HENLEMK-----SYMKERDNLRRVEETLKLBERDQ-----LKESLOET 1915
 Db 2041 AKRIDVLTAGNDRMAEMKQDMQVQYQARIHELQEKLS-QLDQWGEPAATVSSSLDGD 2099

kd golgin) and its upregulation during oligodendrocyte development.";

[2]

SEQUENCE OF 1-980 AND 991-2231 FROM N.A. AND VARIANTS ARG-61;
SER-280; GLU-293; SER-638; THR-819; THR-829 AND ALA-859.
STRAIN=C57BL/6, and Czech II; TISSUE=Brain, and Breast cancer;
MEDLINE=23886257; PubMed=12477932;
Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
Altschul S.F., Ziegler B., Buetow K.H., Schaefer C.F., Bhat N.K.,
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
Datschenko L., Maruina K., Farmer A.A., Rubin G.M., Hong L.,
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
Brownstein M.J., Usdin T.B., Toshiyuki S., Abramson C.F., Prange C.,
Raha S.S., Loughran J.A., Peters G.J., Carninci P., Prange C.,
Bosak S.A., McEwan P.J., McKernan K.J., Abramson R.D., Mullahy S.J.,
Richards S., Worley K.C., Hale S.E., Garcia A.M., Gunaratne P.H.,
Vallalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
Fahey J., Helton E., Kettner M., Maman A., Rodriguez S., Sanchez A.,
Whiting M., Maman A., Young A.C., Shevchenko Y., Bouffard G.G.,
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
"Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences";
Proc. Natl. Acad. Sci. U.S.A. 99:16999-16903(2002).

-!- FUNCTION: May play a role in vesicular transport from the trans-
Golgi (By similarity).

-!- SUBCELLULAR LOCATION: Cytoplasmic, peripheral membrane protein
associated with the trans-Golgi network.

-!- TISSUE SPECIFICITY: Ubiquitous. Highly expressed in
oligodendrocyte precursors, particularly at a stage just prior to
myelination.

-!- DOMAIN: Extended rod-like protein with coiled-coil domains.

-!- SIMILARITY: Belongs to the golgin family.

-!- SIMILARITY: Contains 1 GRIP domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).

EMBL; AF051357; AAC05573.2; -;
EMBL; BC007485; AAH07485.1; -;
EMBL; BC037641; AAH37641.1; ALT_INIT.
EMBL; BC053000; AAH53000.1; -;
PIR; T14265; T14265.
DR MGD; MGI:1859646; Golga4.
GO; GO:0005794; C:Golgi apparatus; IDA.
InterPro; IPR000237; GRIP domain.
InterPro; IPR000345; M_repeat.
Pfam; PF01465; GRIP; 1.
Pfam; PF02370; M; 12.
PROSITE; PS00913; GRIP; 1.
KW Golgi stack; Coiled coil; Polymorphism.
FT DOMAIN 156 2161
FT DOMAIN 335 1011 COILED COIL (POTENTIAL).
FT DOMAIN 1855 2107 GLU-RICH.
FT DOMAIN 2178 2225 GLU-RICH.
FT GRIP.
FT VARIANT 61 61 T -> R (in strain Czech II).
FT VARIANT 280 280 G -> S (in strain Czech II).
FT VARIANT 293 293 G -> E (in strain Czech II).
FT VARIANT 638 638 G -> S (in strain Czech II).
FT VARIANT 819 819 K -> T (in strain Czech II).
FT VARIANT 829 829 A -> T (in strain Czech II).
FT VARIANT 859 859 R -> S (in strain Czech II).
FT VARIANT 2065 2065 R -> S (in strain Czech II).
FT VARIANT 2238 2238 AA; 2257562 MW; 494EA2C11F0165B CRC64;
SEQUENCE

Query Match 7.2%; Score 956.5; DB 1; Length 2238;
 Best Local Similarity 20.9%; Pred. No. 1e-18;
 Matches 515; Conservative 433; Mismatches 867; Indels 589; Gaps 99;

QY 166 DLTEVVYVSEMALK---WITGKES-RHYGET-----KMNQSSRSHTIFRM-ILES-- 213
 DB 48 DVTNENASTQATKSPDGVSKDESSPSQSGDTQTFAKQLQLRVPVSMESLPRSPKESLF 107
 QY 214 REKGPENCGSVKVS---HNLVLDLAGSERRAQTGAAGVRLKGCNINRSFILGQVIK 270
 DB 108 RSSKEP-----LVRTSRESINQDLDCSAAAFDPDPSMESRAEDAPWN----- 151
 QY 271 KLSGQVGGFINVRDSKLTILQNSLGNPKTRIICTITPVSFDTLTALQFASTAKYMK 330
 DB 152 --SDGLSREQLQLRLRMERSLSYRG-----KYSLVTAFTQLQREK--K 193
 QY 331 NTPVYNEVSTDEALLKRYRKEIMDLKQLEVSLETRTAQAMEKDQALALEKD----- 384
 DB 194 KLGILLSQSDKSL---RRISELR---BELOMDQQAQKHLODFDACLXKDYISVL 245
 QY 385 -----LLQKVON-----EKIENLRMLVTSSSLTLOQELK 414
 DB 246 QTVSLKLQRLQNGPMNVDPKPLPGLQAEVHGDTKEKMGVGPVGGTSAKTLEMLQ 305
 QY 415 AKRRVTVCLGKINKMKNKNYADQFNIPNTITTKHLSINLRIDEISVCSSESVFNS 474
 DB 306 QRVKQRE---NLLORCKE-----TIGSHKEQCALLSKEALQELDERLQ 348
 QY 475 TDLTSEIWNPAKLLNQ-----ENISE-----LNSLRADYNLV--- 511
 DB 349 ELKMKELHMAEKTKLITQLRDAKNLIEQLQDQKGVITETKQMLETLEKEDEIAQLR 408
 QY 512 LDYQLRTEKEMELKUKENDLDEFFALER-----TKKQDQEMQLIHEI 556
 DB 409 SHIKMTTQGBELR-EQKESERAFAFELEKALSTAQKTEPAQRRMKWEMDEQKAVERA 467
 QY 557 SNLKNLVHRYNQDLNELSSKVELLREKEDQIKKLOEYIDQKLNIMKOLSYLSIES 616
 DB 468 SEERLRLOHEL---SRVQRAASMAK--KNSEEQVAALQK-LHABELASKEQELSRLEA 522
 QY 617 IEDPKQMKQTLFADATVALDAKRESAFLR-----SENLEKQKMKELATTYQWENDI 669
 DB 523 RE-----RELQEQMRALE-KSRSEYKLITQEKQESLAELELOKAILTSEKNL 575
 QY 670 QLYQSLQAKKQMDVLEKELQSAFNEITKLTSLIDGKVPKDLLCNLELE-----GKITD 724
 DB 576 QELGQAEAYRTRILETSLSEKSLQE-SKTQS-----EHLAVHLEAKKNKHELTA 627
 QY 725 LQKELNKEVEENEARREYVILLSKSLPSE---VERLRKEIQDKSEELHIITSEKDKL 780
 DB 628 LAEQHRTVEGLQO-QODSLWTERLQSLSQHQAAVEELREKYQOEKDAL---LKEKESL 683
 QY 781 FSEVVHKSERVQGLLEIGKTDLDATTQSN---YKSTDQBFQNFKTLHMD---FEQKY 833
 DB 684 FQ-AHIQDMNEKTLEKDKQKMELESSELSEALRARDQALABELSVLRGADKMKQAL 741
 QY 834 KAVLENERMNGEIVNLSKEAKFD--SSLGALKTELSYKTOBLQKRETEVORLNEMEQ 891
 DB 742 EAELEECRRHOREVSGISSEQQELTVRAEKALKDBLSRLGALLDERDEHLRERQARVD 801
 QY 892 LKQLENRDSPLQTVBEREKTILTEKQTLLEEVKTLITQEKDKLQLOESLQIERDQJESD 951
 DB 802 LEHLQKS-----AGELOALAKLDDLHSEQSAAREQAGAYEEOALQMOQK 847
 QY 952 IHDVTNNIDTQOLNLAESLKHQHTINTILSKISIEEVSRLNHEMNTGETKDFEFOOK 1011
 DB 848 V-----LDLETERSL-----LTKQVVMETHKKHVCEELEDAQRAQVQQLERORSELEEK 896
 QY 1012 MVGIDKQDLKAKNTQTLTADVKNDNEIEQORIFSLIOBKNELOQMLQVLESVIAKQKQLT 1071
 DB 897 VRSIAQLODSQLRKNS-TVEKEQARQSLMEKENIILQMRFEQAKIEILIKTLSSKEBSIS 955
 QY 1072 DLKENIEWTIENQELRLGLDELKKQOEIVAQEKNHAIKKEGELSRCTDLRAEVEEKLKE 1131

DB 956 ILHEEVETFKNOEK-----RM-----BKIKQ 977
 QY 1132 KSQOQOEKQOQOQLLNVQOEMSEMOKKINEIENKNEKNTLEHMETERLELAQ----- 1186
 DB 978 KAKEMQETKKLLDQAKLKK-----ELENTVLELSQKE---KQNAQILEWAQANSAG 1028
 QY 1187 -----KLINENY-EEVKSIT-KERKVKELQKSFETERDHLRGYIRIEIATGQTKTEE-- 1236
 DB 1029 ISDTSVRSLEENORQOIESUTGAHQKRLDDVIEAWEKLSQAAELDKHAEQMEKEQGL 1088
 QY 1237 -----LKTARIHLKEHOETIDELRSVS-----EKTAQIINTODLEKSH 1276
 DB 1089 GELRQKVRIVQSEKEELTKVEARLKEAVSQDVALAGLOQLQKQSAVIVS---LSRES 1145
 QY 1277 KLOEETPVHREBELPNVKKVSETQETMNELELLTEQSTTKDSTTLTARTEMERLRINEK 1336
 DB 1146 QLOQSVKEL--EADLGCSEKLSLOEELAKLLADKSOLRVSELTGQVQAAMEKELQSC 1203
 QY 1337 FQESQBEIISLTKERDNLTKIKEALEVKHDQKHEIRETL-AKIQE---SQSQEQOSLN 1391
 DB 1204 KSLHELKSKLEDKSLNKLKSLBELASQDSRCERTKALLEAKTINELVCTSRDKADAILA 1263
 QY 1392 MKEKDNETTKIVSE-----MEQFKPKDSALLRIEIMGL-----SKLOESHDWMS 1439
 DB 1264 RLSQORHTATVGEALLRRMGQVSELBAQLTQLTQTEORTLKSSFQVNTQLEKEKOLKT 1323
 QY 1440 VAKEDDILQLOEVQVQSESQOLKE--NIKEIVAGHLETEELKVAHCLCKQBEETINELR 1497
 DB 1324 MKADIEGLITKEALQOEGGQOQAASEKSCITQLKKE-----LAENINAVTLR 1374
 QY 1498 VNSEKETETSTIOQLEATINDKIQKIOEYKEEQINIKQISEVOENVNELKQFKBEHR 1557
 DB 1375 EELSEKSEHTASLQKSLDGAQLESLSLSP-----SDKAEAISALS--KQH- 1418
 QY 1558 KAKDSALQSIBSKMELTNRLQESQOEIQLIMKX-----EEMKVOQAL 1602
 DB 1419 -----BEQELQLOLAQLOELSLKVDALSKEKMSALEQVDHWSNKFSEMKKQASRL 1468
 QY 1603 QIERDQAKENYKEIVAKKESQEKYQFLKMTAVNETQKCEIEHLKEQFETQKLNLEN 1662
 DB 1469 AQHOSITIKDQAOLDVKAATDAREKEEQCLLKEDLDONK--KPECLKGEVRSKMEK 1526
 QY 1663 -----IETENRILTOILHENLEEMRSVTKERDRLSRVEETLKVERDQKLENRETIT 1714
 DB 1527 KECDLTALKQTAARVE-----LED--CVTQKKEVESLNETLK-NYNQORDTEHSGIV 1578
 QY 1715 RDLEKQOEL-KIVHMLKHEOFTIDKLRGIVSEKNEISNMQKDLSEHNDALKAQD--LK 1771
 DB 1579 QRLQHLBELGEKONKVRABETVLRLREHVSSLEAEELGTVKKELEHVNSSVKSRDGLK 1638
 QY 1772 -IQEELRI---AHMHLKQOETIDKLRGIVSEKTKLSNMOKDLSENAKLAQEKIQE-LK 1826
 DB 1639 ALEDKLELSAAKVELKRAE-----OKIAAIRKQL---LSQMEKQRYAK 1682
 QY 1827 ANEHQILITLKQDVNETQKVSMEQLKKQIKD-----QSLTSLKLEINLNL 1873
 DB 1683 DTENRLELSAQLKREKQVHSLQKLESLKKNLESSPHEVPAVSKSMQSVASP-EQAPDS 1741
 QY 1874 AQELHLENLEEMKSVMKERN-----LRRVEE-----TLKLERDQKESQETKARDL 1920
 DB 1742 QDCTHACKERLCMLQRRLLSEKELRLRLEQGEARPSQPEAQHALLSGKLDCTARQL 1801
 QY 1921 -----ELQOELKTARMLSEKHE-----TVDKLREKISEKTI--- 1952
 DB 1802 EDHVLGICLPEELEKXMKCSLIVSQPMGEETGNNTGVKNWASVDSVQKTLQKEKTCQ 1861
 QY 1953 ----QISDQKOLDKSKDELQKKIQBELQK-----ELQLLRVKED---VNM 1991
 DB 1862 ALBQKVELESIDLVRGAHRLVEKILTLYKESQSQSQEQMDGENCKCVELDERPEENSQ 1921
 QY 1992 SHK---KINMEQOLKQFEPNYLCKBMDNFQITKHLSELEIRIVAKE-RDELRRIKE 2047

Db 1922 SHEZQSVGVTDGLRDLB-SKLTGAERDKQKLSKEVARLQKELRALRREHQBELDKR 1980

Qy 2048 SLKWRDQFIATREIARDRONHOVKE---KLLSDGQOHLMSREKCSRIKELKR 2104

Db 1981 ECEQEAR-----EKLQSQEDLELXHTSLKQLMREFNTQLAQKQOELERTVQETIDK 2033

Qy 2105 YSE-----MDDHYECINRLSLDI-EKEIEPHRIMKKLYVLSVYVTKIKESQHCINKFEM 2158

Db 2034 AQEVEAELSHQOEETQOLHRKIAEKEDDLRTARRVEEILD----- 2075

Qy 2159 DFIDEVEKOKELIKIQLHQQDCVPSRELRLDKLNQMDLHIEILKDFSESEFPISKT 2218

Db 2076 -----ARREEMTKVTDLQ-----TQLEELQKKYQORLEQESTKD--SVTILELOT 2120

Qy 2219 EFOQ-----VLSNRKEMTQFLEEMLNTFRDIEKLNKGIQKENDRICQV-----NNFFNN 2267

Db 2121 QLAQKTTLISDSKLKEQLRE-----QVNELEDRKRYEKNAACATVGTGYKGNLYHT 2174

Qy 2268 RIATINVESTFEF-----ERSATISKESWQDLKSLKEKNEKLFKNYQT----- 2310

Db 2175 E-VSLFGEPTFEYLKVMFEYMMGRETMTAKVITTVLKFPDQQAQKILEREDARLMSW 2233

Qy 2311 LKTS 2314

Db 2234 LRTS 2237

RESULT 9

RBP2 PLAVB

ID RBP2 PLAVB STANDARD; PRT; 2867 AA.

AC Q00799; Q9N2M3; 2867 AA.

DT 01-APR-1993 (Rel. 25, Created)

DT 15-MAR-2004 (Rel. 43, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Reticulocyte binding protein 2 precursor (PVRBP-2).

GN RBP-2 OR RBP2.

OS Plasmodium vivax (strain Belem).

OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.

OX NCBI_TaxID=31273;

RN [1]

RP SEQUENCE FROM N.A., AND REVISIONS TO 2438-2439.

RX MEDLINE=20299192; PubMed=1617731;

RA Galinski M.R., Xu M., Barnwell J.W.;

RT "Plasmodium vivax reticulocyte binding protein-2 (PVRBP-2) shares structural features with PVRBP-1 and the Plasmodium yoelii 235 kDa rhoptry protein family.";

RT Mol. Biochem. Parasitol. 108:257-262(2000).

RL [2]

RP SEQUENCE OF 1189-2439 FROM N.A.

RX MEDLINE=92315338; PubMed=1617731;

RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;

RT "A reticulocyte-binding protein complex of Plasmodium vivax merozoites.";

RT Cell 69:1213-1226(1992).

CC -!- FUNCTION: Involved in reticulocyte adhesion. Specifically binds to human reticulocyte cells.

CC -!- SUBCELLULAR LOCATION: Membrane-bound (Probable).

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/ebis-sib.ch/announcement/> or send an email to license@ebi-sib.ch).

CC

CC EMBL; AF184623; AAF76525.1; -.

DR HSPF; F03069; IGCM.

KW Malaria; Receptor; Signal; Transmembrane; Repeat.

FT SIGNAL 1 21 POTENTIAL.

FT CHAIN 22 2867 RETICULOCYTE BINDING PROTEIN 2.

FT DOMAIN 22 2805 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 2806 2826 POTENTIAL.

FT DOMAIN 2827 2867 CYTOPLASMIC (POTENTIAL).

FT DOMAIN 44 133 ASN-RICH.

FT DOMAIN 560 758 LYS-RICH.

FT DOMAIN 1112 1285 LYS-RICH.

FT DOMAIN 2758 2785 7 X 4 AA TANDEM REPEATS OF H-D-D-T.

FT REPEAT 2758 2761 1.

FT REPEAT 2762 2765 2.

FT REPEAT 2766 2769 3.

FT REPEAT 2770 2773 4.

FT REPEAT 2774 2777 5.

FT REPEAT 2778 2781 6.

FT REPEAT 2782 2785 7.

SQ SEQUENCE 2867 AA; 331433 MW; 6E7D8CA71AFBFD3 CRC64;

Query Match 7.2%; Score 956.5; DB 1; Length 2867;

Best Local Similarity 21.2%; Pred. No. 1.2e-18;

Matches 567; Conservative 486; Mismatches 932; Indels 693; Gaps 119;

Qy 329 MKQTPYVNEVST--DEALLKRYKKEIMDLKKQLEFVSLETRAQAMEKDLQAL----- 379

Db 248 LENPQYNDISKYDEK-VKEYKKIEDMQICLKDSYIKNFKAIMGANLKNLALNGIYI 306

Qy 380 -----LEEKDLQKVQNEKIEINLTMLVTSSSLTI-QQELKAK--- 416

Db 307 HWWYLTCSSTKYDDIVKEYAIHINDEKKSIFMDNMKK--IHKSAIDTLKKQKALNT 364

Qy 417 --RKRVTWLCKINQK-----KNSNYAQDF--NIFT-----NITTKT----- 450

Db 365 SLDSKRTETIIGIEMIKFNLHLTKIRYASAFATKSIPLQKVESDIYRVELTKLFFVAA 424

Qy 451 -----HKLINLREIDESVCSDESDFNTLTLSEIENWPAKTLNQINSEINSLRA 505

Db 425 KHYADPKFSLEHLKMPENLSKSKKMLYSTFFKLEGLDKINKINTLMGSSQSDLTSLIA 484

Qy 506 DYDNLVDLYEQL-RTEKEEMELKKEKNDLDFEAL-ERKTKKQOEQMLIHISLKN-- 561

Db 485 DSEKIKKSAESLINSSEI-----AKYALDSNEKINEIKKYDQNIILKRVREFINKSGL 539

Qy 562 --LVKHREYVNOQLENLSSKV-ELLREKEDQIKKQLEVIDSKLENTKMDLSYSLESTE 618

Db 540 ITSVMGTSQLESQDIQETKIEIKKKKDLERGERKEFINIMN-EIKKKKKSNSNST 598

Qy 619 DPQMKQTLFDAETVALDAKESAFRLSENLEKMKELATYKQMDIOLYQSOLEA 678

Db 599 NSEFTDKLELETFEGNLTVMKGYLOEIEDIKVKNEDRSKQIIEOHLK-YTSDNRD 657

Qy 679 KKKQVLDLEKELQSAFNEITKLTSLIDGKVPKDLCLNLEEGKITDLOKELNKEVEENA 738

Db 658 NVKTLISKNDIEIKYIEKIEKLN--DAPSKD-----KFTTEKTNLQNKVKKIIDEFH- 709

Qy 739 LREEV-ILLSEKSLPSEVERLRKEIQ--DKSEELHIITSEK-----DK 779

Db 710 -KEDLQLLNLSLSKFYEHEQKLYNEASTIEIKDLHQTKKEYEKLERKMSFNQIILDK 768

Qy 780 LFSEVVKESRVQGLLE-IGTK-----DGLATQSNYKSTDDQEPQNPKT 824

Db 769 LNTLNDNLKNEKINVEQNTYINKVMSDSLTNLTAEYDNLRSALDGYRADETELTKYN 828

Qy 825 LHMDFEQKYKMWLEBENE-----REKLTITEKLOQTELEVKTUT-GEKDDLK 852

Db 829 RINERKEKFLSTLKEQEDDIPDGKNIYEYNNHKNVWVNEKHKISSDINQCNENIIEK 888

Qy 853 EAQKFDSSLGALKTELSTYKTELQLE-----KTREVOERLNEMEQ-----LKEQL 896

Db 889 NLETFTNLVQTLDAHTGKQKQKVHDLQKFTNLKLNLELESFGKSLNGSASTYTKQI 948

Qy 897 ENRDSPLQTVB-----REKLTITEKLOQTELEVKTUT-GEKDDLK 935

Db 949 ENIRKNIDTTLNFAKNSSSEKSLSENITKNKADLIKKLDQHTQIEKHFTFENEEMS 1008

Qy 936 QLOESLQIERDQLKSDIHDVTWNNDTQELRNLAKESLKHQETINTLKSISSEVRNL 995

Db 1009 PLSVVKKEKNRVESDMS-----ELIKQNTKINAILLEYNNKSDRFN 1052


```

Db 909 FAEIETELQAEVMEQ-----QHCKVLYLLSQLOOSQMAE----- 945
QY 935 KQLOESQIERTDOLKSDIHDVTNNMIDTQBLRANALESLKQHQTINTLKSIS--EYS 992
Db 946 KQLEESVSEKEQQLS-----TLKCODELEKREVCENQOQLRENEIIRKQKULLQVAS 1001
QY 993 RNLHMENT-----GETKDFEQKQWGI-DKQDLEAKNTQTLTADVK 1034
Db 1002 ROKHLPKDTLLSPDSFEYVPPKPSRVKKEFLQSDIEDLYKCSHVSNEHEDGDG 1061
QY 1035 DNEIIEQORXIFSLIQ 1050
Db 1062 DDEGDDEWPTKLK 1077

RESULT 11
KF4A MOUSE
ID KF4A MOUSE STANDARD; PRT; 1231 AA.
AC P33174;1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 41, Last annotation update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chromosome-associated kinesin KIF4A (Chromokinesin).
GN KIF4A OR KIF4 OR KNS4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, AND TISSUE
RP SPECIFICITY.
RC STRAIN=ICR; TISSUE=Brain;
RX MEDLINE=95014709; PubMed=7929562;
RA Sekine Y., Okada Y., Noda Y., Kondo S., Aizawa H., Takemura R.,
RA Hirokawa N.;
RT "A novel microtubule-based motor protein (KIF4) for organelle
RT transports, whose expression is regulated developmentally.";
RL J. Cell Biol. 127:187-201(1994).
RN [2]
RP SEQUENCE OF 91-240 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=93077686; PubMed=1447303;
RA Aizawa H., Sekine Y., Takemura R., Zhang Z., Nangaku M.,
RA Hirokawa N.;
RT "Kinesin family in murine central nervous system.";
RC CC -!- FUNCTION: REQUIRED FOR MITOTIC CHROMOSOMAL POSITIONING AND BIPOLAR
CC SPINDLE STABILIZATION.
CC CC -!- SUBCELLULAR LOCATION: Nuclear. Associated with mitotic
CC chromosomes.
CC CC -!- TISSUE SPECIFICITY: EXPRESSED IN PYRAMIDAL CELLS IN JUVENILE
CC HIPPOCAMPUS, GRANULAR CELLS IN JUVENILE CEREBELLAR CORTEX AND IN
CC ADULT SPLEEN.
CC CC -!- SIMILARITY: Belongs to the kinesin-like protein family.
CC Chromokinesin subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D12646; BAA02167.1; -.
CC PIR; A54803; A54803.
CC HSP; P17119; 3KAR.
CC MGD; MGI:108389; Kif4.
CC InterPro; IPR001752; kinesin_motor.
CC Pfam; PF00225; kinesin; 1.
CC PRINTS; PR00380; KINESINHEAVY.
CC SMART; SM00129; K1Sc; 1.

```

```

DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
DR PROSITE; PS0067; KINESIN MOTOR DOMAIN2; 1.
KW Motor protein; Microtubule; ATP-binding; DNA-binding;
FT Nuclear protein; Coiled coil. KINESIN-MOTOR.
FT DOMAIN 1 350 COILED COIL (BY SIMILARITY).
FT DOMAIN 1001 1231 GLOBULAR.
FT NP_BIND 88 95 ATP (POTENTIAL).
FT FT_CONFLICT 112 112 I -> S (IN REF. 2).
SQ SEQUENCE 1231 AA; 139551 MW; F34F2C2D21158FE4 CRC64;

Query Match 7.1%; Score 952.5; DB 1; Length 1231;
Best Local Similarity 26.0%; Pred. No. 7.5e-19;
Matches 367; Conservative 223; Mismatches 491; Indels 329; Gaps 51;

QY 7 VAVCVVRPLNSREESLG-ETAQVYKTDNNVIYQVDGSKSFNDRVFGHGNETTKNYEE 65
Db 10 VRVALRCPLVSKIEKGCOTCLSFVPGEQV--VGNDKSFYDFVDFSTEGEEFNT 67
QY 66 IAAPIIDSAIQGYNGTIFAYGQTASGKTYTMMGS-----EDH---LGVIPRAIHDFQIK 118
Db 68 AVAPLKGVPFGYNATVLAAYGQTGSGKTYSGGAYTAEQEHDSAGIVIPRVIQLLFXEIN 127
QY 119 KFPDREFLLRVSYMEIYNETITDLCGT-QMKPLIIREDVNRNVYVADLTVEEVYVISEM 177
Db 128 KKSDFEFTLVSYLEIYNEBILDLCSREKATQINIREDPKEGKIVGTETKTVLVSAD 187
QY 178 ALKMITGKSRHYGETKMNQSRSHITFRMILESREKGEPSNCEGSKVSHLNLVLA 237
Db 188 TVSCLEQGNNSRTVASTAMNSQSSRSHAIPTISIEQRK-----NDKSSSFPSKLHLVLA 243
QY 238 GSREAAOTGAAGVRLKGCNINRSLFILGOVKKLSGQGVGFINYRDSKLTIRLQNSLG 297
Db 244 GSERQKTKAEGDRREGINRINRGLLCLGNVISALGDDKGNFVPRDYSKLTIRLQNSLG 303
QY 298 GNPRTIITCTTPV--SPDETTLTALQFASAKYMNTPYNEVSTDEALLKRYKEIMDL 355
Db 304 GNSHTLMACVSPADSNLEETLNLTVADRARKINKPIIN-----IDPQAAELNHL 355
QY 356 KQLEEVSL-----ETRAQAMEKDQLAQLLEKDLQKQVQNEKIEIN 396
Db 356 KQVQQLQLLLOAHGGTLPGDINVPSENIQSLMERKQ--SLVEENKLSRGLSEAAQ 413
QY 397 LTRML---VTSSSLTQQLKAKRRVTCWLGKINK---MKNSYADQFNIPNTITK 449
Db 414 TAQMLERILITQEAQNEKNAKLEIRRHAAKVDLQKLVTLEDQELKENIICNLQOV 473
QY 450 THKLS-----INLREIDESVCS-----BSDFVSNLTDLTSE-----LEWNP 486
Db 474 TAQLSDEAAACMTATIDTAGADTQVQSSPDTSRSDVFS-TQHALRQAMSKELIENK 532
QY 487 A-----TKLNQENIESELSRADYD---NLVDYEQRLTEKEEMELKIK-EKNDL 534
Db 533 ALALKEALAKKNTQ--DNQLQPIQFQODNTKNESEVLSIQREKEVLVLELQAKKDA 590
QY 535 DEFEALERTKDKQEMOLITHEISNKLNVKREVNQDLLENELSSKVELLRKEKDIKKL 594
Db 591 NQAKLSERRRKLQEL--GQIALKXKLQ-----EQSKLLKKESTEHTVSKL 637
QY 595 QBYIDSQKLENIKMDLSYSLESIEDPKQKQTLFPAETVAL---DAKRSAPLRSE-NLE 650
Db 638 NQELRMKQNRVOL-MRQMKEDAEPKQWQKQ-KDEVQLKDERDKRQYELKLRNFQ 695
QY 651 -----LKEQKELATTYKOMENDIQLYQSLQLEAKKQVQDLQLEKLSQAFNEITKLSLID 705
Db 696 KQSNVLRRTKEEAAAANKLQALQ-----KQEAKEKETSQSGNESTAARM---744
QY 706 GKVPKDLLCNLELEGTITDLQKELNKEVENEALREEVILLSELKSLPSEVRLKEIQD 765
Db 745 ----KWLGN-----EIVVSTEEAKRHLNGLLEERKILAQDVAQL-KEKRE 787
QY 766 KSEELHITSEKDKLFSVVKHESRVQGLLEELIGTKYKDDLATTSQSNYSKTDQEFQNFKL 825

```



```
QY 522 REMELKKEKNDLDEFEALERTKDKOEMQLIHEISNLKLVKRE-----VYNQDLE 574
Db 593 NQAKLSERRRKLROELGQWTELK-----KLGEQSKLLRSTKTVAKMKEIQ 644
QY 575 NELSSVLLREKEDQIKLQEIYDSQKLNIRKDLSSIESIEDPKMKQTLFDATVA 634
Db 645 GMDQVRQLMRQKDEAKFRTWKQKTEVIQI-----KEKDKRQVYELL----- 690
QY 635 LDARSAFARSENLEKKEKELATYKQKENDIQLYQSLAKKQKQVLDLEKELQSAF 694
Db 691 ---KLERDFQKQANV-LRRKTEEAASANKELQKQKAMEKRDQSO---SKGMGAA 743
QY 695 NEITKLTSLDGVKPDLLCN-LELECKITDLOKELNKEVEENALREEVILLSELKSLP 753
Db 744 SRV-----KNWLANEVELVSTEAQHNLNLDLEDKILQAQDI---AQLKQKT 788
QY 754 SEVERLRKEIQKSEELHIITSEKDLFSEVHKESRVQGLLEIGTKKDDLATQSNYK 813
Db 789 DAGERIPTKIRRT-----YTVAELENLEEEASVTK----- 819
QY 814 STDQEFQNFKLHMDPFQKQKWLVEENRMQEIWNLSKEAQKFDSSIGALKTKELSYKTQ 873
Db 820 -----QLESLEMEELRSA 833
QY 874 ELQKTRVQERLNMELQKLENRDPSPLQTVREKTLIPEKIQOTLEEVK---TLTQE 930
Db 834 QIADLQKLLDADGEEENVKRWETISNIMEAKALKVITELVSSKVGSKLESSVKON 893
QY 931 KDDLKQKQESQIERDQKLSIDHTVNMNIDTQBLRNALSLKQKQHTINTILSKISE- 989
Db 894 RAHVADLQKNIFEERNQ-----AEMETEHSQSL---MQLEHQHKEKILYLLSQLOK 943
QY 990 EVSNRLNMEENTGETKDFQKQV-----GIDKKQDLKAKNTQTLTADVKNELIEQ 1041
Db 944 QASVPVTIEELFAEITEREQKMERLKFQDEIEKMKALCEKQKQLL---QENDMTYK 999
QY 1042 QKTLFSLQKNEQLQOMLESVIAEKEQLKTD-----LKENIEMTIENQEBLR 1089
Db 1000 KLALHVASGK-----LHNILPAEICSPPSPDFIPPKRGKRTNAKSAVILEDL- 1053
QY 1089 LLGDELKQKQEIQAQKNAHKKEGELSRCTDRLAEEVEKL---KKSQOLQKQKQILLNV 1146
Db 1054 LSESESEESDDKQWEPGNNSQSKLTKSCCKARCGNMGCRKTKQ-----NC 1104
QY 1147 QEEMEMQKKEINENLKNELKXKELTLEHMETERLELAQKLNENYEVKS----- 1197
Db 1105 SDDCFCDPSCKRNDRNMDGKHEDQSL-SENSKID-----YPDVTAGGSFTTTPC 1155
QY 1198 ITKEREKVLKQKSFETERDHLRGVIREATEATGLQTKBELKIAHILKHEQHTIDELRRS 1257
Db 1156 VPTTKVLKREISD-----IGQVLSIKLQKPKSTASASVMESQEN---QTS 1199
QY 1258 VSEKTAQII 1266
Db 1200 ILTKKKKVL 1208

RESULT 13
KF4A_CHICK
ID KF4A CHICK STANDARD; PRT; 1225 AA.
AC Q90640; Q90608;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chromosome-associated kinesin KIF4A (Chromokinesin).
GN KIF4A.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
```

```
RP SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, AND TISSUE
RP SPECIFICITY.
RP STRAIN=White leghorn; TISSUE=Embryonic retina;
RX MEDLINE=95181533; PubMed=7876303;
RA Wang S.Z., Adler R.;
RT "Chromokinesin: a DNA-binding, kinesin-like nuclear protein.";
RL J. Cell Biol. 128:761-768 (1995).
RN [2]
RP SEQUENCE OF 728-1088 FROM N.A.
RP STRAIN=White leghorn; TISSUE=Embryonic retina;
RX MEDLINE=94151328; PubMed=8108415;
RA Wang S.Z., Adler R.;
RT "A developmentally regulated basic-leucine zipper-like gene and its
expression in embryonic retina and lens.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:1351-1355 (1994).
CC -1- FUNCTION: REQUIRED FOR MITOTIC CHROMOSOMAL POSITIONING AND BIPOLAR
SPINDLE STABILIZATION.
CC -1- SUBCELLULAR LOCATION: Nuclear. Associated with mitotic
chromosomes.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN PROLIFERATING CELLS;
CC -1- NEUROEPITHELIUM OF EMBRYOS.
CC -1- SIMILARITY: Belongs to the kinesin-like protein family.
Chromokinesin subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U18309; AAC59666.1; -.
DR EMBL; U04821; AAA18960.1; -.
PIR; A56514; A56514.
HSSP; P17119; 3KAR.
DR InterPro; IPR001752; kinesin_motor.
DR Pfam; PF00225; kinesin; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
DR PROSITE; PS0067; KINESIN MOTOR DOMAIN2; 1.
KW Motor protein; Microtubule; ATP-binding; DNA-binding;
KW Nuclear protein; Coiled coil.
FT DOMAIN 1 351 KINESIN-MOTOR.
FT DOMAIN 352 1003 COILED COIL (BY SIMILARITY).
FT DOMAIN 1004 1225 GLOBULAR.
FT NP BIND 88 95 ATP (POTENTIAL).
FT CONFLICT 1087 1088 KG -> RI (IN REF. 2).
SQ SEQUENCE 1225 AA; 138923 MW; FA01ED83425F5875 CRC64;

Query Match 6.9%; Score 917.5; DB 1; Length 1225;
Best Local Similarity 25.2%; Pred. No. 6.2e-18; Indels 349; Gaps 54;
Matches 344; Conservative 242; Mismatches 428;

QY 3 EGVAVCVVRVPLNSREESLG-ETAQVYVKKTDNNVYQVVGSKSFNEDRVFHNGETKN 61
Db 6 EKGIVRVVRCRPLVPKTESCCQWCLSPVGPQVI--VGSDKAFTYDVEDPSVEQE 63
QY 62 VYEEIAAPLIDSAIQYNGTIFAYQTASGKTYTMWG-----SEDH---LGVIPRAHDIF 114
Db 64 VENTAVAPLIRGIFKGYNATVLAYQTGSGKTYSMGTYTASQEHDPMSMGVPRVILKLF 123
QY 115 QKIKKPPDREFLLRVSYMEIYNTITDLCGT-QKMKPLIREDVNRNVVADLTVEVY 173
Db 124 KEKEQFQDWFEVLKVSYLEIYNEIDILLCSSRERSQSISREDPKGKIKVGLTERNVA 183
QY 174 TSEMALKWLTKEKSRHYGETKMNORSRSRSHITFRMLRSREKGEPSNCGSVKVLHNL 233
Db 184 SARDTVSCLEQGNCRVTASTANSSQSSSHALFTICIDQKKK---NDKNSFSFKLHL 239
QY 234 VDLAGSERAQOTGAAGVRLKEGCNINRSFILGQVTKKL-SDQVGVGFNYRSLKTRIL 292
```

Db 240 VDLASERQKTKAEGDRLEKGINNGLCLGNVISALGEENKKGFPYRDSKLTLL 299
QY 293 QNSLGNPKTRIICTIPP--SFDETLTALQFASTAKYMKNTPVNEVSTDEALLKRYK 350
Db 300 QDSLGGNSHTLMTACVSPADSNLEETLNTLYRADRKIKKPIVN-VDFQAEHLNHLKQ 358
QY 351 EIMDLKKQLEE-----VSELETRAQ-----MEKDLAOLLEEKDLLOK-----VQNEK 393
Db 359 QVOOLQVLLQAHGGTLPVINSMAPSENLSLMEKNQ--SLMBENEKLSRGLSEAGQT 416
QY 394 IENLTMLVTSLSLTQLQELKARKKRVWCLGKINKMKNSYADQNPITNTTTHKL 453
Db 417 AQWLEKLIIVTEQ--NEKNAKLEQLQAHVCKLDLQ----- 452
QY 454 SINLREI-DESVCSBDSVFNSTDLTSEIE-----WNPATKLKLNQENIESELNSLRAD 506
Db 453 ---LLETVEDEELKENVEVRNQQVLAQFQSESAAAAEAATEMANAE----- 497
QY 507 YDNVLVDYQLRTEKEEMELKLEKNDLDFEALERKTKKQDMQMLIHEISNKNIVKHR 566
Db 498 -----QDAAGEAETGVTRKSSDDFTTQHAL-RQAQMSKELVELNKALALKEALAKK 548
QY 567 EYVNOQLENSKVELLEKEKDIKKLOEYIDSKLENIKMDLSYLSIESIEDPKMKQT 626
Db 549 MIQN-----DSQLEPIQ-----SOYQTNIK-DLELEVSNLQ----- 578
QY 627 LPDAETVALDRAKRESAFLRSNLEKEMKELATTKYKQEMNDIOLYQSOLEAKKQMVDL 686
Db 579 -----KEK-EELILASMAKQVN--QAKLSERR----- 605
QY 687 EKELQAFNEITKLSLIDCKVPKLLCNLEEGKITDLOKELNKEVEENEALREEVILL 746
Db 606 -RLQ-----ELEGQINELKKNL-----EQAKLL 629
QY 747 SLKSLPSEVERLRKEIODSEELHIITSEKDKLPSEVVHKSRSVQGLLEIGTKYDULA 806
Db 630 KKESTERTVSKLNGEIREM-----KNQVQ-LMRQ----- 660
QY 807 TTQSNYKSTDOFQNFKTHMDFEQYKQV--LEENERNQ--EIVNLSKEAQKQPSLSICA 863
Db 661 -----KEDAEKFRQWK-----QKQKEVILQKEDRKQVYELKLEDFQK-QASVLR 707
QY 864 LKTELS-----YKQBLQKTRVQRNLNE--MEQLKEQLENRDSPLQTVREK 910
Db 708 RKTBEAANAANKRLDALQKREADRKESQNRGMGVAARVKSMLANEVEVLVSTEAR 767
QY 911 TLITEKQQLTEEVKLTQEKDOLKQLESQIERDQKLSHDHVTNNIDTQEQRLNAL 970
Db 768 ---RHLADLLEDKRLAQELLQLEKESGENPPSKLRRTYSITDIQASEMDL----- 818
QY 971 ESKLQHQBTINTLKSITSEVSRNLHMBENTGETKDFQKQVGVHDKKQDLEAKNTQTLT 1030
Db 819 -SLSKQIESLET-----EMELRSQAQIAD-LQKLLDAD----- 849
QY 1031 ADVKDNEIIEOQKIFSLIQKNEHLQOMLESVIAEKEQLKTDLKENIEMTIENQBEALL 1090
Db 850 ---NGDRVKQKQWMDNIATILEAKCALKYLLGELVSSKVQ-ESKLESSLQSKTNCSDI--- 902
QY 1091 GDELKKQOEIVAOEKNAHKEGELSRTCDBLAEEVEKLEK-----SQOLEKQOQLNV 1146
Db 903 -----QKMLIERNHATEWEAFQ---NOLLQEQHQOQEVLYLLSQFOQKAPKGV 952
QY 1147 QEEMSEMQKINEIENLKNLKNELTLHEMETERLELAQKLENIEVEKSIITERKVLK 1206
Db 953 EDSLSEQEKQOE-----RLKFEKELEKVR-----EICEKNQELQE-NDVLKQKMLLV 1001
QY 1207 ELQKSFETERHLRGYREIATGLQYKEELKIAHIHL-----KEHQETIDELRRSVSEK 1261
Db 1002 QVASGQKLRDQ-----QVSPSPDQPFQVIPPKPKTRQTVAKPRAPTE- 1047
QY 1262 TAQINTQDLKSHTKLQEEIPVHLHERQELLPNVKYKSETQET 1304
Db 1048 ---MNVEELFSDSESGEE-----EDAEMWVP-VKAAKGTCKS 1080

RESULT 14
SPOF SCHPO
ID SPOF SCHPO STANDARD; PRT; 1957 AA.
AC Q10411; Q9USE9;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Sporulation-specific protein 15.
GN SPO15 OR SPAC1F3.06C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
SEQUENCE FROM N.A., FUNCTION, SUBUNIT, AND SUBCELLULAR LOCATION.
RP MEDLINE=20107136; PubMed=10639340;
RA Ikemoto S., Nakamura T., Kubo M., Shimoda C.;
RT "S. pombe sporulation-specific coiled-coil protein Spo15p is localized
to the spindle pole body and essential for its modification.";
RL J. Cell Sci. 113:545-554(2000).
RN [2]
SEQUENCE FROM N.A.
RP MEDLINE=972;
RC STRAIN=21848401; PubMed=11859360;
RA Wood V., William R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgourou J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Felwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Mottier S., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Welter J., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritze C., Holzer H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
RN [3]
SEQUENCE OF 705-871 FROM N.A.
RP STRAIN=968 h30;
RC STRAIN=968 h30; PubMed=10759889;
RA MEDLINE=20223868; PubMed=10759889;
RA Ding D.-Q., Tomita Y., Yamamoto A., Chikashige Y., Haraguchi T.,
RA Hiraoka Y.;
RT "Large-scale screening of intracellular protein localization in living
fission yeast cells by the use of a GFP-fusion genomic DNA library.";
RL Genes Cells 5:169-190(2000).
CC -!- FUNCTION: Has a role in the initiation of spore membrane
formation.
CC -!- SUBUNIT: Monomer.
CC -!- SUBCELLULAR LOCATION: Spindle pole body.
CC -!- SIMILARITY: Belongs to the MPC70 family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch.

CC EMBL; Z70690; CAA94624.1; --
 DR EMBL; AB027811; BAA87115.1; --
 DR PIR; T38077; T38077.
 DR GenDB SPombe; SPAC1F3.06c; --
 KW Sporulation; Coiled coil.
 FT DOMAIN 199 785 COILED COIL (POTENTIAL).
 FT DOMAIN 804 1235 COILED COIL (POTENTIAL).
 FT DOMAIN 1320 1471 COILED COIL (POTENTIAL).
 FT DOMAIN 1481 1723 COILED COIL (POTENTIAL).
 SQ SEQUENCE 1957 AA; 222785 MW; 3F480CA06171D9DA CRC64;
 Query Match 6.7%; Score 893.5; DB 1; Length 1957;
 Best Local Similarity 20.3%; Pred. No. 4e-17;
 Matches 425; Conservative 448; Mismatches 800; Indels 419; Gaps 76;
 212 ESREKGPSCGSKVSHNLVDLAGSERRAAQTGAAGVRLKGCNINRSLFILGQVIKK 271
 62 DSSEELKROEVGRHRSIDSIDAKLGSSEGS-TASSALPLTPRSPNAGWLLVR----- 115
 272 LSDQGVGFNYRDSKUTRILONSLGNPKTRICTTTPVSFDTLTALQFASAKYMKN 331
 116 -----GGLLDSPILDINSVTQKSNLLNELKQVASKLAALHEHNGIILSLSSNKKDKN 169
 332 TPVNEVSTDEALLKRYK-----ETMDLKKOLEEVS--LETRAQAMEKDQ- 375
 170 TSSVTTLTSEEDVSFYOKFLTMNSNFSKASEAYDLRSOLLITVEKLDKKEKDYKKE 229
 376 -----LAQLLEKOLLQVQNEKIENLRLMLVTSSSL--TLQQ-----ELKAKRRVTVCL 425
 230 DVSIKASLAEEQASNSKIRGEQ-ERLEKLVSSNKTVSTLRQENSIRABCK----- 281
 426 GKINKMKNVADQFNPTNITTKHKSIN-----LLREIDESVCSSESVFSTLPTLS 480
 282 -----TLQEKLEKCAINEEDSKLEELKHNVANYSDAIVHK-DKLI 321
 481 ELENPATKLANQENIESELSRADYDNL-----VLDVEQLRTEKEEMELKKE 530
 322 E-----DLSTRISEFNLKSERDTLSIKNEKLEKLRLNTIGSLKDSKTSNQSLEEEVVELKE 378
 531 KNDLDEFBALRKKKQOEMQLIHEISNLKLVHREYVNOQDLENELSKVELLREKEDQ 590
 379 SNRTHSOLTAESKLSSEFQ--ENKSLKSGSI-----DEYQNNLSKDKMVKQVSSQ 428
 591 IKKQYEYI--DSQKLENIMDLSYSLESIEDPKQMKQTLFPAETVALDAKESAFRSEN 648
 429 LEEARSLAHATGKLAIEINSERDFQNKIKDFEKIEODL-----RACLNSSS 475
 649 LELEKMKELATYKQMDIQLYQSOLEAKKMQVDLEKELQSAFNEITKLTSLIDGKV 708
 476 NELKEK-----SALIDKQOELANLRQIKQKKVSESTQSSLOSLODI-----LNEKK 525
 709 PKDLNLELEKITTDLQKELNKEVEENALREEVILLSELKSLPSEVRLKKEITQDKSE 768
 526 KHEV-----YESQLNELKGELOTEISNSELH-----SOLSTLAAEKEAAVATNNELSE 574
 769 ELHIITSEKDKLFPSEVHVESRVQGLLEIGIKTKDDLTATQSNYKSTDOEFQNFKTLMD 828
 575 -----SKNSI-----QTLNAPQOEKLAISVMQKENQNFSSLDTSFKKLUNESHQ 620
 829 FEOYKMWLEENRMNOEIVNLSKEAQFDSISGALKTELSYKTOBLOKREKREVOERLNE 888
 621 LENNHQITTKLQDTSKSLQOQLERANFEQ-----ESTLSDENNDRTKLLKLEESNKS 676
 889 MEQLKEQLENRDSPLQVREKTLITEKIQOULEEVKTLTQEKDDLKQLESQLEPQOL 948
 677 LIKKQEDVDSLEKNIQTLKEDLRKSEEARFSLKLEARNLRVIDNLRKGHETLEAQNDL 736
 949 KSDIHTVNNITDQQLNALRESLKHQETINTLASKISEEVSRLNHEENTGETKDEF 1008
 737 HSSLSDAKNTNAILSLSELTKSSSDVXRLTANVETL--TQDSKAMKQSFSLVNSYQISNL 795

QY 1009 QOKMVGIDKKQDLEAKNTOTLTADVK--DNEIIEQORKIFSLIOEKNELOOMLESVIAE 1065
 DB 796 YHEL--RDDHVNMQSQNTLLSESKLTKYDCEMLTOQN--WILI---DNVQKLHKHVNQ 848
 QY 1066 KEOLKTDLKE--NIEMTIENOEELRLGDELAKQOQEI-VAQEKNHAIKKEGELSRCDRL 1122
 DB 849 ESKV-SELKEVNGKLSLDLKN-----LSSSLNVAISDNDQILTQALAESKNYDLS 897
 QY 1123 AEVEEKLKESQOOLQKQOQLNVQEMSEMOKKIN---EENLKNELKNKELTLEHMET 1179
 DB 898 EQESAQLNSGLKSLAEKQLLHTENBELHRLDKLTKLIEESKSDGLKLTARQOEI 957
 QY 1180 ERLELAQKLENVEEVKSIKERKVLKE-LQKSFETERD--HLRGYRREIETAG--LQT 1233
 DB 958 SNLK-----EENMSQQAITSVKSKLDETLSKSKLEADIEHLKNKVSVEVERNALLAS 1012
 QY 1234 KEBLKIATHLKEHQETIDELRRSVSEKTAQIINTQDLEKSHTKLQEBEIPVLHBEQE--L 1291
 DB 1013 NERLM--DDLKNGENIASLQTEIEKRAE-----NDDLOKLSVSVSEYENLL 1059
 QY 1292 LPNVKVSQETQWNELELLTEOSTTK--DSTTLARIEMERL-----RLNEKQESOEI 1344
 DB 1060 LISSQTNKSLDKTNQLKYI-EKNVQKLDEKQORNVLEBELTSKYKIGREENAQIKDEL 1118
 QY 1345 KSLTKER-----DNLKTIKEALEVKHQKEHIRETLAKIOESQSKOQSINMK 1393
 DB 1119 LALRKSKKHQDLCANFVDDLKESDALE---QLTNEKNELVLSQSNSNEA---LV 1171
 QY 1394 EKONETTKIVSEMQ-FPKPDSALLRIEIMLGLSKLOESHDMMKSVAKEDDLO-RLQ 1451
 DB 1172 EERSDLANRLSDMKKSLSDSDNVISRLVRV-----NDELDTLKDKDSLSQTS 1224
 QY 1452 EVLQSESDQLKENTKEIVAKHLETEBELKVAHCCLEKEQETINELRVNLSEKETEISTIQ 1511
 DB 1225 EVCQDRDLDLS-----LKGESEFNKYAVSLRELCTK----- 1257
 QY 1512 KQLEAINDKQNKIQEIVEKEEQNLKIQISEVQE--NVNELQKFKEHRAKQDSALQSIRESK 1570
 DB 1258 -----SEIDVPVSEIIDDNFVFNAGNFSLSRLTVLSLENYLD-----AFNQVNFK 1303
 QY 1571 MLELTNRLOESQEBEIQIMIKEEMKRVQEAQIQRDQLEKQENTKEIVAKMESQEKEVQF 1630
 DB 1304 KQELNRLTTTDAFTKVVADEKLQEHDDWLIQDGLD-----KALKDSEKNF 1353
 QY 1631 LKMTAVNTEQEMCEIEHLKEQFQETQKLENIENTENIRLTQILHENLEEMRSVTKERDD 1690
 DB 1354 LRKEA--EMTENIHSLEEGKEETKEIAELSSRLSDNQLATNKNQDLHL-----NOE 1405
 QY 1691 LRSVEETLKVREDQLENIRLITRDLKQEBELKIVHMHLEKHOETIDKLRGIVSEKTN 1750
 DB 1406 IRLKEDVLK-EKESLITISLESLSNORQESSL-----LDKXNE 1443
 QY 1751 ISNMQKOLEHNSNDALKAQDLKIQEBELRIAHMLKEQQETIDKLRGIVSEKTKDNKQMD 1810
 DB 1444 LEHMLDDTSRKNSSLMEXIESINSSDDKSPELASAVEKLGALQKLHSESISLMEINISQ 1503
 QY 1811 LENSNAKQ--EKIQELKANEHQILITLKDVNETQKVVSEMEQLKKQDKQSLTLKLE 1867
 DB 1504 LQEAKEKIQVDESTIQEL---DHEITASK--NNTYEGKND-----KD----- 1540
 QY 1868 TENINLAQELHENLEEMKSVM-KERDNLRRVETLKLERDQLEKESIQETKARDLEIQOEL 1926
 DB 1541 ----SIRDLSENIEQLNNLLAEESKAVKLSSEKESLQFNLSRLADLEYHKSQVSESL 1596
 QY 1927 KTWARM--LSKEHKEITVDKIREKISEKTIQISDIQKLDKSKDELOKKIQELQKQELQL 1983
 DB 1597 GRSKLKLAATTEELQLAENERLSLITRMLDQNVQKDLQVQKDLQVQKDLQVQKDLQV 1656
 QY 1984 RVKEDVNMHKKINEMEQKQFEPNYLCKCEMDNFQTLKKLHSELEERIVAKERDE-- 2041
 DB 1657 SLQKECKI---KSNTVESLQ-----DVLTSVQARNAEDEVSRSVDKIR---RRDRC 1704

QY 2042 -----LRRIKESLKMEDQFI-ATLREMIARDQNHQVKEPKRLSLDGOQHLMSLREK 2094
 Db 1705 EHLGKLLKHLQSHQEQHETFFRAEQQMTQLGFLKVTYKQEKI-----LKKL 1753
 QY 2095 CSRIKELLKRYSEMDHYECNRLSLDLKX-----IEPHRMKKLVLSVYTK 2144
 Db 1754 NLKQEQILPSSIL--VYESYIR-----DIEKEIIVLQERLNGHLSQQLPKGYGYFFKT- 1807
 QY 2145 IKEOHECINPFMDFIDEVEKO-----KELLIKIHLQOQCDVPSRE 2187
 Db 1808 -----NRVEMVLDSPKQVAKIQLFLAGREFFVKEFKEDLEKCAABEEK 1850

RESULT 15
 MLPL_YEAST
 ID MLPL_YEAST STANDARD; PRT; 1875 AA.
 AC Q02455;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Myosin-like protein MLPL1.
 GN MLPL1 OR YKR095W OR YKR415.
 OS Saccharomyces cerevisiae (Baker's Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288c;
 RA MEDLINE=93247549; PubMed=8483450;
 RX Koelling R., Nguyen T., Chen E.Y., Botstein D.;
 RT "A new Yeast gene with a myosin-like heptad repeat structure.";
 RL Mol. Gen. Genet. 237:359-369 (1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94205265; PubMed=8154186;
 RA Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,
 RA Renacha M., Jimenez A., del Rey F., Ballasta J.P.G., Revuelta J.L.;
 RT "The complete sequence of a 15,820 bp segment of Saccharomyces
 cerevisiae chromosome XI contains the UBI2 and MPL1 genes and three
 new open reading frames.";
 RL Yeast 9:1349-1354 (1993).
 CC -!- FUNCTION: Myosin-like protein that is probably involved in DNA
 CC repair.
 CC -!- SIMILARITY: SOME, TO THE TPR ONCOGENE.
 CC -!- CAUTION: Ref.2 misquotes the gene name as "MPL1".
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR ENBL; L01992; AAA34783.1; -;
 DR ENBL; X73541; CAA51948.1; -;
 DR ENBL; Z28320; CAA82174.1; -;
 DR PIR; S38173; S38173.
 DR Germline; 140074; -;
 DR SGD; S0001803; MLPL1.
 DR GO; GO:0005635; C:nuclear membrane; IDA.
 DR GO; GO:0005654; C:nucleoplasm; IDA.
 DR GO; GO:0006606; P:protein-nucleus import; IDA.
 DR KW Coiled coil; DNA repair.
 FT DOMAIN 69 487 COILED COIL (POTENTIAL).
 FT DOMAIN 531 1678 COILED COIL (POTENTIAL).
 FT DOMAIN 1834 1866 COILED COIL (POTENTIAL).
 FT CONFLICT 301 301 R -> A (IN REF. 1).
 SQ SEQUENCE 1875 AA; 218455 MW; 683A0D34C9066867 CRC64;

Query Match 6.7%; Score 891.5; DB 1; Length 1875;
 Best Local Similarity 20.7%; Pred. No. 4.4e-17;

Matches 469; Conservative 467; Mismatches 756; Indels 579; Gaps 96;
 QY 462 DEVCSESDFVNTLDTLSHIEWNPATKLNQENIESELSNLRADYDNLVDLYEQLRTE- 520
 Db 18 DERLNAIAFFGCSLQVKSFD-GDVVHLLNDKLL--QFNELKSENKLVTVSDFELKASS 74
 QY 521 -----KEEMELKLEKNDLDEFALERTKTKQDQEMQLIHEISNLKLVKHREVNQDL 573
 Db 75 LKIDGLKTEMENVIRENDKI-----RKERND-----TFVKFESVENEKM 114
 QY 574 ENELSKVLLREKEDQIKKQVYDSOKLENKMDLSYSLESIE-----DPKQ 622
 Db 115 --KLSSELEFVRKLDLDEEKKETQSNQOORTUKI-LDERLKEILVRVNNNSNSECK 171
 QY 623 MKQTLFDAET-----VALDAK-RESAFLRSNLELKEKMK--ELATTYKQ- 664
 Db 172 LRSTMIDLETQGYITNDLNGSRTELETKTQELTLQSNNDWLEKLSKNEQVLSYRQK 231
 QY 665 -----MENDIQLYQSLEAKKQVQVLEKLSQAFNITKLTSLIDGKVPK 710
 Db 232 TDKVILDIRNELNRLNFQMERITNDVLKKNNELSKSLQEKLLKLGSLDSLSNE-KQ 290
 QY 711 DLLCNLELGKITD-LOKELNKEVEENEALRE---EVLILSELKSLPSEVERLRKIQK 766
 Db 291 EFSAEWSLKQRLVLDLLESQUNAVKELNSIRELNTAKVIADDSKKQTPENEDLLKELQ 350
 QY 767 SEELHIITSEKDKLFEVVHVESRVQGLLEETGKTDDIATQSN-----YKST 815
 Db 351 KEKL--AQCEKEL-----RLSSITDEADENENLSAKSSDDFLFKQLIKERT 399
 QY 816 DQEFON-FKTLHMDPEQYKVVLENERMNOQIVNLSKBAQKFDSSLGALKTELSTYKQ 874
 Db 400 KEHLQQLTETFIVELEHKVPIINSFKRTDM-----LENELNAAALL 441
 QY 875 LQEKTRVQERLNEMEQLEOLENEDSDPLQTVREKTLITKQLQTLSEVTKLTQEKDDL 934
 Db 442 LEHTSNEKNAVKELNAXKQKLVCEQDLQTLTKQRLDLQCIQVLLITNSVNSDKGL 501
 QY 935 KQLQESLQIERDQKSDIHTVNMNIDTQEURNALESKQHOETINTLKSISB--SVS 992
 Db 502 R--XEEIQINIMQED--DSTITESDSQKVVTRELVEFK---NIIQLEKNAELLKV 553
 QY 993 RNLHEENTGETKDFQOQKVGIDKQDLKAKNTQTLTADVKDNEIIFQQRKIFSLIOEK 1052
 Db 554 RNL-----ADKLESKEKSKQSL-QKIESETVNEAKEALITLSEK 593
 QY 1053 NLOQMLESVIAEKEQLKTDLKENIEMTIENQEBELRLGDELKKQOEIVAEKNAIKKE 1112
 Db 594 MDLESRIEELQKELEELT-----SVNED-----ASYSNVTIK-- 627
 QY 1113 GELSRTCDRLAEVEBEKLEKSKQQLQEQOQQLNVQ--EEMSEMOKKINEIENLKNELK-- 1168
 Db 628 -----QLTETKRDLESQVQDLQTRISQITRESTENMNSLLNKEIQDLYDSKSDISL 679
 QY 1169 NKELTLEHMETERLE-LAQKLNENYEVKSIK-----ERKVLKELQKSFTEHDHL--R 1220
 Db 680 GKEXSRIABERFKLSNTSLDITKAENDQLRKRFQVLTQNTILKQDSKTHETLNEYVSK 739
 QY 1221 GYTRIEATGLQTKELKIAHILKHOETIDELRRSVSEKTA---QIINTODLEKSHYK 1277
 Db 740 SKLSIVETELLNLKEBQKL-RVHLERKWLK--QELNLSPEKDSLRINMVTQLQTLQKRED 796
 QY 1278 LOBEIPVLHBEQELLPNVKVSETQETMNELELLTQSTTKDSTTLARIEMERLRNEKF 1337
 Db 797 LLEBTRKSCQ-----KKIDELEALSEL-----KXETSQKDHIIKQLEDN---NSNI 841
 QY 1338 QESQEEIKSLTKERDNLKITKEALEVHKHOLKEHIRETLAKIQESQSKQBSQSLNMEKON 1397
 Db 842 EWTQNKIEALKDYEVSVITSDSKQTDIEKLYQVKVKSLEKEIEEDKIRLHTYNNMDETIN 901
 QY 1398 ETTKIVSEMQFKPKDSALLRIETIEMGLSKRLQESHDEKMSVAKKDDLQRIQEVLOSE 1457
 Db 902 D-----DS---LRKELEKSKIN--LTDAYSQIK---EYKDYETTSQSILQQT 940

Qy	2458	IRLLREMLRESQAQDTSVLSSEH-TPQPSNKLTCGGSGIVQNTKALILKSEHRLK	2511
Db	1786	---LTSNPQAQDSSNRNVQSEEDTEKKKEGPVK-RGEAIEQTKS-----NKRPID	1834
Qy	2517	EISKLKQONEQLIKQKNELLSNNQHLISNEVKTWKERTLKEEAHKQVTCENS	2567
Db	1835	EVGELKNDKEDDTPTTENINE-----SKTKTDEEE--EKETVK-VNDENS	1874

Search completed: July 29, 2004, 09:36:12
Job time : 52.5784 secs

QY	1458	SDQLKENIKEIVAKHLETEEBELKVAHCCQLKQDETINELRWNLSEKET-----BISTIQKQ	1513
Db	941	NSKLDSEPKDFT-----NQIK-NLUTDETSLDEKLSLKEQ	975
QY	1514	LEAINDKLQNKQOEIYEKEEQNLNIQISEVOENVNELKQFKHEKAKDSALQSTESKMLE	1573
Db	976	MFNLNNELDQ-KKGMEKEKADFKRISILQNNKNEVEAKSEYSEKSLQNDLDOOTI	1034
QY	1574	LTNRLQESQEEIIMIKEEEMKRVQZ-----ALQIERDQLKENTKEIVAKM	1620
Db	1035	YANTAQNVE--QELQKHADVSKTISELREQLHTYKGQVKTLNLRDQLENALKE-----	1087
QY	1621	KESQEKYEYFLKMTAVNETQEKMCIEHLKEQFETQKLNLENITENIRLTQIILHENLEE	1680
Db	1088	---NEKSSWSQK-----ESLLEQLDLSNRSIRIDLSQN---KLHYDQIQI	1126
QY	1681	MRSVTKERD-----LRSEVETLKVVERDOLKENIRETIT-RDLQEK-QEELKIVHMHKE	1732
Db	1127	YTAADKEVNNSWNGPLNILITLRERDIL--DTKVTVAERDACKMLFQKISLMDVELOQ	1184
QY	1733	HQETIDKLRGIVSEKTEISNNQKD---LEHNS--DALKAQDLKIQBELRTAHMHLKEQ	1787
Db	1185	ARTKLDNSR--VEKENHSSIIQOHDIMEKLNQLNLLRESNITLRNELENNNNKKCELQ	1241
QY	1788	ETIDKLRGIVSEKTKLSNNQKDLENSNAKLOEKTQELKANEBHQLITLKQDWNQTKKVS	1847
Db	1242	SELDKUK-----QNVAPIESELTALKYSMEKQOELK-----LAKEEVHRWKESQ	1287
QY	1848	EMEQLKQIKQDQSLTSLKLEIENLNAQELHENLEEMKSVMKERDN-LRVEE---TLK	1902
Db	1288	DI--LEKHEQLSSDYEKLESEIENKEEL-ENKERQGAEGAEKEFNLRROQERLKTSK	1344
QY	1903	LERDOLKE-----SLOETKARDLEIQ-----QELKTARMLSKKHETV	1940
Db	1345	LSQDSLTEQVNSLRDAKNVLNLSLSANARIEBLONAKVAQGNQLEATRKQOEDREKAS	1404
QY	1941	DKLREKISEKTIQ---ISDIQDLKSDKDELQKTIQELQKKEQLQLLRKVEDNNSHKKI	1996
Db	1405	RELQAKLEESTSYESTINGLNEIITLKEIEKQIQIQ--QLQATSANEQNDLS---	1458
QY	1997	NEMEQLAKQEPFNLYLCKCENDFNQLTQKGLHESLEETIRIVAKERDELRRIKESLKMERDOF	2056
Db	1459	NIVESMKKSF-----EDKIKFKTEKIQ-EVANEK	1486
QY	2057	IATLRMTARDORQHGVKPEKRLSDGOQHLMSLEBKCSRIKELLKRYSEMDHDHYECIN	2116
Db	1487	ILEAQERLNQPSNINMEIIRKKHSEHEQSVQKIRE---ABEALUK-----	1530
QY	2117	RLSIDLEKEIEFHRIMKKLVLSYVTKIIBEQHECINKEPMDPIDVEFKQKELLIKIOH	2176
Db	1531	RIRLPTBEKI-----NKIIRKKE--ELEKEFEKVEE-----RIKS	1565
QY	2177	LQOQCDVPSRELRDLKNQNDMLHIEILDKDFSESEPPSIKTFEPQVLSNRKEMTOFLRE	2236
Db	1566	MEQSGEI-----DVLFRKOLEAKVOEKQEL-ENEY-----NKK-----LOE	1601
QY	2237	WLNTFRDIEKLKNGITQKENDRICQVNNFFNNRIIATMNESTETEPEERSATISKEWEOQLKS	2296
Db	1602	ELK---DVPHSHSIDDERDKI-----RAEIESR---LRREFNELQ	1638
QY	2297	LKEKNEKLFKNYQTLTKTSLASGAQVNPFTQDNKNPHVTSRATQTTTBEKIRELNSLHEAK	2356
Db	1639	IKKKSDFDEKQOAMMKTLLE-----RKLAKMESQISETK	1673
QY	2357	ESAMHESKILIKMOKELEVTNDI IAKLOAKVHESNKC-----LEKTKETIOVLQKVALGA	2412
Db	1674	QSAEGPPKSVNNVQPL-----IQLPRKTEENSNSFPNLLSGEKLKLKNSKSSGGF	1726
QY	2413	KQYKBEIEDLKMVLKIDILEKMKNAKEPEKEI-----SATKATVEYQKEV	2457
Db	1727	NBFTSESDNKTITONNDKRESLANKTDPPTHLPSFNIPASRGLISSSTLSITDNDDEE-	1785

GenCore version 5.1.6
 Copyright (c) 1993 - 2004 Compugen Ltd.
 OM protein - protein search, using sw model
 Run on: July 29, 2004, 09:39:47 ; Search time 133.008 Seconds
 (without alignments)
 6280.361 Million cell updates/sec

Title: US-10-045-631B-88
 Perfect score: 13329
 Sequence: 1 MAREGAVCVVRPLNSR.....SQPMPHASSGKDVPCKTQ 2663

Scoring table: BLOSUM62
 Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000
 Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : Published Applications_AA.*
 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				SUMMARIES	
Result No.	Score	Query Match	Length DB ID	Description	
1	1234	9.3	1382	16	US-10-437-963-176714
2	1183	8.9	240	10	US-09-893-519A-67
3	1125.5	8.4	3225	16	US-10-408-765A-254
4	1077.5	8.1	1388	12	US-10-332-089-2
5	1077.5	8.1	1388	14	US-10-146-473-82
6	1077.5	8.1	1388	15	US-10-173-999-32
7	1077.5	8.1	1388	16	US-10-188-832-164
8	1064	8.0	3899	14	US-10-171-311-4
9	1064	8.0	3917	14	US-10-171-311-8
10	1058	7.9	3907	14	US-10-171-311-2
11	1058	7.9	3925	14	US-10-171-311-6
12	1055	7.9	3911	15	US-10-370-685-100
13	1055	7.9	3911	16	US-10-408-765A-1839
14	1052.5	7.9	3878	14	US-10-080-608A-11
15	1041.5	7.8	2383	14	US-10-082-830-260

16	956.5	7.2	1232	14	US-10-116-712-670	Sequence 670, App
17	956	7.2	1232	16	US-10-408-765A-2153	Sequence 2153, App
18	956	7.2	1235	15	US-10-334-143-8	Sequence 8, Appli
19	944.5	7.1	1232	14	US-10-116-712-664	Sequence 664, App
20	944.5	7.1	1232	14	US-10-116-712-669	Sequence 669, App
21	944	7.1	1237	15	US-10-334-143-33	Sequence 33, Appl
22	938.5	7.0	694	12	US-10-425-114-59725	Sequence 59725, A
23	910.5	6.8	1401	16	US-10-287-226-142	Sequence 142, App
24	899.5	6.7	1392	16	US-10-287-226-86	Sequence 86, Appl
25	898.5	6.7	1392	16	US-10-473-574-27	Sequence 142, App
26	893.5	6.7	1957	15	US-10-369-493-2070	Sequence 86, Appl
27	891.5	6.7	1875	15	US-10-369-493-22285	Sequence 27, Appl
28	886.5	6.7	1905	15	US-10-369-493-22285	Sequence 22285, A
29	873.5	6.6	1827	15	US-10-369-493-5368	Sequence 5368, App
30	870	6.5	10203	16	US-10-661-809-23	Sequence 23, Appl
31	867.5	6.5	963	14	US-10-080-608A-20	Sequence 20, Appl
32	867.5	6.5	963	15	US-10-370-685-109	Sequence 109, App
33	866.5	6.5	957	14	US-10-240-145-145	Sequence 145, App
34	866.5	6.5	957	14	US-10-080-608A-16	Sequence 16, Appl
35	866.5	6.5	956	14	US-10-370-685-105	Sequence 105, App
36	863	6.5	956	14	US-10-080-608A-17	Sequence 17, Appl
37	863	6.5	956	14	US-10-370-685-106	Sequence 106, App
38	856	6.4	967	15	US-10-080-608A-21	Sequence 21, Appl
39	856	6.4	967	15	US-10-370-685-110	Sequence 110, App
40	854.5	6.4	1979	14	US-10-205-823-419	Sequence 419, App
41	854	6.4	1805	10	US-09-820-843A-73	Sequence 73, Appl
42	854	6.4	1805	12	US-10-282-122A-63513	Sequence 63513, A
43	850.5	6.4	1394	16	US-10-287-226-84	Sequence 84, Appl
44	848.5	6.4	963	14	US-10-080-608A-22	Sequence 22, Appl
45	848.5	6.4	963	15	US-10-370-685-111	Sequence 111, App

ALIGNMENTS

RESULT 1

US-10-437-963-176714
 ; Sequence 176714, Application US/10437963
 ; Publication NO. US20040123343A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa, Thomas J.
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Wu, Wei
 ; APPLICANT: Boukharov, Andrei A.
 ; APPLICANT: Barbazuk, Brad
 ; APPLICANT: Li, Ping
 ; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(53221)B
 ; CURRENT APPLICATION NUMBER: US/10/437,963
 ; CURRENT FILING DATE: 2003-05-14
 ; NUMBER OF SEQ ID NOS: 204966
 ; SEQ ID NO 176714
 ; LENGTH: 1382
 ; TYPE: PRT
 ; ORGANISM: Oryza sativa
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: PAT_MRT4530_74437C.1.pep
 US-10-437-963-176714

Query Match 9.3%; Score 1234; DB 16; Length 1382;
 Best Local Similarity 28.6%; Pred. No. 1.5e-53;
 Matches 432; Conservative 254; Mismatches 514; Indels 310; Gaps 54;

Qy	7	VAVCVVRPLNSRESIGETAGQVYWKTDNNVI-YQVDGSKSFNDFRVHGNKTKNVYEE	65
Db	4	IHVAVRPLTA--EDAGSP--WVSGNAIALSTQPSIRFEDRIFGECRTADYGA	58
Qy	66	IAAPIIDSAIQVNGTIFAYGOTASGTYTMCSEDLHGVIPRAIHDFOKIKKFPDPREF	125
Db	59	RTKHIIVDSVRGNGTVFAYGQTNKGTMYRSGNPGIPLAVHDLFTTIEHLDREF	118

QY 1094 LKQOEIVAEKHNHAKKEGELSRTCDRLAEVEEKLKESQOQKQOQLL-----1144
 Db 1098 RSRFGCVLGVGRGRTPRGGVDGWVACVAPNSSAAQWSSAGGEVAGVLRACAGGIWH 1157
 QY 1145 --NQOEEMSEKQKINEINELNKNELNKLLEHM-----E 1178
 Db 1158 LVASVADTMFGDSEWSEKALLGLPLMLATATPSGIVHLVEGVAIGALVOLHKGFLRTFEE 1217
 QY 1179 TERLELAQK-----LNEN--YEEVKSIKER-KVKLEOKSFETERDLRGYIREIEATG 1230
 Db 1218 SERLLAKERNIDTGVNENELHQOLLSITEERDKLSEIK-----YMSVINESE-- 1266
 QY 1231 LQTKELKIAHILKHEOETIDELRRSVSEKTAQIINTQDL-----EKS 1274
 Db 1267 -----LIQAKATIDELSSRISIVEAKMKNYWLPLSNLIDECFDSAYNKE 1313
 QY 1275 HTKLOEIPVIEHEQELLPNVKVSQTQETMNELELLETE---OSTTKDSTTLARIEMERL 1331
 Db 1314 NTKLRMQIRWQPELD-----AHRGRLEKAEINEMKLMMDTKYLKASTKLLKOLSPYCREVL 1368
 QY 1332 RLNEKFOESQ 1341
 Db 1369 RLKEQLKESQ 1378

RESULT 2

US-09-893-519A-67
 ; Sequence 67, Application US/09893519A
 ; Publication No. US20030027243A1
 ; GENERAL INFORMATION:
 ; APPLICANT: ANADYS PHARMACEUTICALS, INC.
 ; APPLICANT: THOMPSON, Craig
 ; APPLICANT: MOORE, Jeffrey
 ; APPLICANT: BUURMAN, Ed T.
 ; APPLICANT: BRADLEY, John
 ; APPLICANT: DESILVA, Thamara
 ; APPLICANT: HARRIS, Sandra
 ; APPLICANT: KOMARNITSKY, Svetlana
 ; APPLICANT: MENDILLO, Marc
 ; APPLICANT: MOORE, Daniel
 ; APPLICANT: MCCOY, Melissa
 ; APPLICANT: SANDERSON, Karen
 ; APPLICANT: HAQ, Tariq
 ; APPLICANT: ZHU, Shuhao
 ; APPLICANT: LONG, Fan
 ; APPLICANT: DAVIDOV, Eugene
 ; TITLE OF INVENTION: ANTIFUNGAL COMPOUNDS AND METHODS OF USE
 ; FILE REFERENCE: 0342/10548-US2
 ; CURRENT APPLICATION NUMBER: US/09/893,519A
 ; CURRENT FILING DATE: 2001-06-28
 ; PRIOR APPLICATION NUMBER: US 60/215,164
 ; PRIOR FILING DATE: 2000-06-29
 ; PRIOR APPLICATION NUMBER: US 60/224,457
 ; PRIOR FILING DATE: 2000-08-10
 ; NUMBER OF SEQ ID NOS: 146
 ; SOFTWARE: Patentin version 3.1
 ; SEQ ID NO 67
 ; LENGTH: 240
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; OTHER INFORMATION: Corresponds to SEQ ID NO: 140
 ; PUBLICATION INFORMATION:
 ; DATABASE accession number: Human Genbank/CAA78727
 ; DATABASE entry date: 1993-01-10
 ; RELEVANT RESIDUES: (1)..(240)
 ; US-09-893-519A-67

Query Match 8.9% Score 1183; DB 10; Length 240;
 Best local similarity 100.0%; Pred. No. 6,1e-52;
 Matches 240; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 LLRVSMYIYNETITDILCTQKMKELIREDVNRNVYVADLTEEVVYTSSEALMKWITKG 185
 Db 119 LLRMSYMEIYNEEINDLL--VPEHRKQLIHESIERGIYVAGLREIEVTCPEQVLEWSPG 176
 QY 186 EKSRRHGETKMNQSRSHITFIMLESREKEGPPS-----NCEGSVKVSHNLVDLAGSER 241
 Db 177 ESRHHIGETNMNVYSRSHITFRWVIESREKVDSEAGESSCD-AVRVSVNLVDLAGSER 235
 QY 242 AAGTGAAGVRLKGCNINRSLFLIGQVKKLSDG--QVGGFINRYSKLTIRILONSLGNN 239
 Db 236 AAKTGAAGVRLKGCNINRSLFLIGQVKKLSDG--QVGGFINRYSKLTIRILONSLGNN 235
 QY 300 PKTRILICTIT--PVSFDEITLALQFASAKYKNTPVNVSVSTDEALLKRYKREIMDLKK 357
 Db 296 ANTAIIINITLAQVHADETSSLOFASRALURVNCACNBIIDALLKQREIEHLRA 355
 QY 358 QL-----EVSLETRAQAKEDQALQLEKDLQKQVNEKIEINLFRMLVTSLSLTLQ 411
 Db 356 KLRSELEKESISLE-----LEEEKAKQKQRLIE--QAKKIENL-----SSLVINS 401
 QY 412 E-----LKAQRKRVVTCWLGKINMKNSVADOFNPTNITTKTKLSINL---LREID 462
 Db 402 ERDDRTTVSSKNKRLTWCPLGRSLQDQGVLESQVBDPPSVTVRGRNMEMPLHFEELI 461
 QY 463 ESVCSSES-----DVFS-----NTLDTLSE-----481
 Db 462 QSCSESISKHVTDAYSSGSLSCEDDSLPDSHALLHVTSRRKPNMTMKSDQQLMGLASER 521
 QY 482 -----JEWNPATKLLQNIENESINSLRADYDNLVDYEQLRTEKEMEL-KLKERNDL 534
 Db 522 IIPQELNDKWTYQ--SQENIKACVNGLSARESEAILVIKQLEDQIKLLEKESFPQNL 579
 QY 535 DEFALERKTKQDQEMOLIHEISNLKVLKREVNQDLENELSSKVELLREKEDQIKKL 594
 Db 580 D--DVLELAT--QOKASFHE-----KYEELQONALVAQEQAKI--ANEKLSQBA 624
 QY 595 QRYIDS--QKLENIKMDLSYSLESTEDPKQMKOTFLDAETVALDAKRESAFILRSNLEIK 652
 Db 625 YELFTGIFVETESIAVQDQSTRSDNALSPFIEELF-----QNLFWAKNFTVS 673
 QY 653 ERKKE-----LATTY-----KOMENDITQYCSQLEAKKQKQVLDLEKLOAFNHI 697
 Db 674 IDNYOPFCSIYIDLVLTGYGKPGKIGDRTTLAPAGLLLAELKEVEAK---QFVCGDI 730
 QY 698 TKLTSLLID-----GKVPKDLLNLELHGIKIDLOKELNKEVENEALREEVILLSELKSL 752
 Db 731 TQFSVIRDYENISNCLREKLSKLEMEKKILDEQSLDQKD-----ELQRL 775
 QY 753 PSEVERLKEIQDKSEELHIITSEKDKLFSVVHVESRY-----QGLLEEIGTKDDLAT 807
 Db 776 KSLSECEKAMEDCNQNEL-----EKDSIILSELLTQKEVYLSLSSSLMKESIRKELDR 832
 QY 808 TOSNYSKTDQEFONFKTLHMDFEQKYKWLNEE-----RMQEIIVNLSKEAQKFDSSL 861
 Db 833 TTKLKETENKLN-----SIQEKILESEKAEAKQKIKLQSQRTLLERDLRKDSFT 886
 QY 862 GALKTELSYKTOEL-----QKTRVQQLNEMEQIKQLENRDSPLQTVREKTLITEKLQ 918
 Db 887 VDKRHEQSVKSELAGIYQAVQIOEDYKLEHAFDMEABIASLQ--EALVTIAEK-E 943
 QY 919 QTLIEVKTLTOEKDOLKQLESQIERDOLKSDIHDVTNNNIDTQBLRNAME-----SLK 974
 Db 944 EALSRELLTSAVEDLESLNASETSLSLEET-AVLTKLDAESISKKLEASISLS 1002
 QY 975 QHOETINTLKSISIEEYSRNLHMEENTGETYKDFQKQKMGVIGDKQDLKAKNTOTLTADVK 1034
 Db 1003 REKEDMGIELT-----DVLLEWESERSWTAK-----KAYLEAK--QKLNLCNK 1045
 QY 1035 DN-EITTEQQRKIFSLIOEKLELQOMLESVIAEKEQLKTDLENIENTIENQOELRLIGDE 1093
 Db 1046 NNCKLSDELTKV-----QIQRISSGGFRASPHWERLELRRRTAVVKAARGIVGSGPS 1097

QY 1801 TDKLSNMOKLNSNAKLOEKIOELKANEHQIITIKVDVNETQKVKSEMEQKKQIKDQS 1860
 Db 1 TDKLSNMOKLNSNAKLOEKIOELKANEHQIITIKVDVNETQKVKSEMEQKKQIKDQS 60
 QY 1861 LTLKLEIENLNAQELHENLEEMKSVNKRDNLRVVEETLKLBDQIKESLOETKARDL 1920
 Db 61 LTLKLEIENLNAQELHENLEEMKSVNKRDNLRVVEETLKLBDQIKESLOETKARDL 120
 QY 1921 EIOQELKTARMLSKHEKTEVDKREKISEKTIQISDIQKOLDKSDKDELQKKIQELOKKEL 1980
 Db 121 EIOQELKTARMLSKHEKTEVDKREKISEKTIQISDIQKOLDKSDKDELQKKIQELOKKEL 180
 QY 1981 QLLRVKEDVNVSHKKINEMEQKKOFEPNYLCKCEMNFQTLTKKHSLEIRIVAKERD 2040
 Db 181 QLLRVKEDVNVSHKKINEMEQKKOFEPNYLCKCEMNFQTLTKKHSLEIRIVAKERD 240

RESULT 3
 US-10-408-765A-254
 ; Sequence 254, Application US/10408765A
 ; Publication No. US20040101874A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Ghosh, Soumitra S.
 ; APPLICANT: Fahy, Eoin D.
 ; APPLICANT: Zhang, Bing
 ; APPLICANT: Gibson, Bradford W.
 ; APPLICANT: Taylor, Steven W.
 ; APPLICANT: Glenn, Gary M.
 ; APPLICANT: Warnock, Dale E.
 ; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
 ; FILE REFERENCE: 660988.465
 ; CURRENT APPLICATION NUMBER: US/10/408, 765A
 ; CURRENT FILING DATE: 2003-04-04
 ; NUMBER OF SEQ ID NOS: 3077
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 254
 ; LENGTH: 3225
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-408-765A-254

Query Match
 Best Local Similarity 8.4%; Score 1125.5; DB 16; Length 3225;
 Matches 599; Conservative 529; Mismatches 963; Indels 718; Gaps 119;

QY 251 RLKGCNINRSLFILGQVVKILSDQGVGGFI---NYRDSKLTRI-----L 292
 Db 680 KAKEISLNQ---LIEBFKNADNNSAFTALSERDQLLSQVKELSMVTETRAQVKOL 735
 QY 293 QNSLGGNPKTRICITITPVSFDETLT---ALQFASATKMYKNTPTVNEV-----STD 341
 Db 736 EMNLAEARQRRLDYESTAHNLLITEQIHSLSIEAKSKVKLEVQLNELDDVQLQFSEQ 795
 QY 342 EALLRVKPEIMDLKKQLEEVSLERAQAMEKDLQALLERKOLLQKQVONKIEINLRML 401
 Db 796 STLRSQSQJONKESEVLEGAERVHRHSSKVELSQALSQKEL-----EITQMDQLL 848
 QY 402 VTSSS---LTLQELKAKKRRVTCWLGKINKMKSNDYADQNPINITYTKHLSINLR 459
 Db 849 LEKKRDVETLQOTTEKQOQVTEISFSTWTEKQVQLN--EEKSLGVFI--KTLKEQLNLS 905
 QY 460 EIDES-----VCSSEDFVNTLDTLSEIPWNPATKLLNQNTESELSNLRADYD-----508
 Db 906 RAEBAKKEQVEDNEVSGLKQNYDEM--SPAG--ISKELQHFLLKQNEQKRRKLQ 962
 QY 509 -----NVLVDYQLKTE--KEEMELKKRKNLDLDEFEALERK-----TKK 546
 Db 963 AALNRKELLQVRSLREELANLNDKSKGTEPLSETERGEVEEDK--ENKEYSEKCVTSK 1020
 QY 547 DQEMQLTHEISNLKNLVKRVNQDLENELSSKV-----ELLREKEDQIK 592
 Db 1021 CQEIEIY-----LKQTISEKEVELQHRKDLLEKLAABEQFQALVKQMNQTLQDKTNQID 1075

QY 593 KLOEYIDS-----OKLENIKMDLSYSLESIEDPKOMKOTLFDFAETVAL-----635
 Db 1076 LLOAEISENQAIIOKLITSNTDAS-----DGDVALVKETVVISPPCT 1118
 QY 636 -----DAKRESAFLSENLLEKEMKELATYTKOMENDIQLYQSOLEAKKQKQVLDLE 687
 Db 1119 GSSHWPEPEEKILALEKEQOLQKLEALTSRKAILKKAQEKERHLREELQKQKDY 1178
 QY 688 KELQSAFNEITK-----LTSLIDGKVP-----709
 Db 1179 NRLQEQFDEQSKENENIGDQLRQLQIQRRESIDGKLPSSTDQOESCSSTPGLEPLFKATE 1238
 QY 710 -----KDLLC-----NLEEGKITDLQKEL 729
 Db 1239 QHHTQPVLESNCPDWPMSHSEDASALQGTTSVAQIKALKEIAEKVELEKSVSTTSSEL 1298
 QY 730 NKEVEENEALREEV-----ILLSELKSLPSEVERLRKEIQKSE-----ELHI 772
 Db 1299 TKKSEEVFQOQINKQGLIEISLKTVSHEAFVHAESLQCKLESSQLIAGLEHLRELQ 1358
 QY 773 ITSEKDKLFSEVVHKBRSVQGLLEIEIGTKDKDLATTQSNYSKSTDQEFQNFKTHLMDFEQK 832
 Db 1359 KLDELQKLISKKEEDVYLSGQLE-----KEAALTKIQTEIIEQE--DLIKALHTQLEMQ 1412
 QY 833 YKMWLEENEMNQEIYNLSKEAQFSDSLGALKTELSYKTOELQEKTEV-----QERLNE 888
 Db 1413 AK---BHDERIKOLQVCELMKQK--PEEIG---EESRAKQOIQKLOAALISRKELKE 1464
 QY 889 MEOLKEQLENRDSPLQTVREKTLITEKIQOTLEEVKTLTOEKDDLKQLESQIBERDQ 948
 Db 1465 NKSQBELSLARG---TIER---LTKSLADVESQVSAQNKEDQTVLGRLLALQERDKL 1517
 QY 949 KSDIHDVANNIDTQOLRNALSKOHOBTINTLTKSKISEEVS---RNLMHMENTGETKD 1006
 Db 1518 ITEMDSILEN---QSLSSCESLKLAEGLTEDKEKLVKEIESLKSKEIAEST-----1568
 QY 1007 EFQOKMVGIDKKODLEAKNTQTLTADVDKNEIIEQOKKIFSLIQEKNEQLQOMLESVIAEK 1066
 Db 1569 EWQEKHKELQKEYEILLQSYENVS---NEAERIQHVVEAVRQEQKQELYGKLRSTEANK 1623
 QY 1067 EOLKTDKLENIEWTIENQOEELLLGDGLKKQOEV--AQKNHAIKKE-----GELSTC--1119
 Db 1624 KETEKQLOEAOEMEMKEMKRFKAK--SKOOKILEBENDRLRAEVHPAGDTAKECNE 1681
 QY 1120 -----DRLARVEEKLKEKSOLOEQKQOLLNVQEMSEMOKKINEIENKNEK---NK 1170
 Db 1682 TLLSSNASMKEELERVKMEYETLSKKFOSLMSKDSLSE-----EVQDLKHQIEGNVSK 1735
 QY 1171 ELTLEHMETERLELAQKLNENYEVKSYTKERKVLKELQKGFETERDHLRGVIRIEATG 1230
 Db 1736 QANLE--ATEKHD--NQTNVTEEGTOSIPGETE-----EODLSLMSR---PTC 1777
 QY 1231 LOTKEELKIAHILKHEQETIDELRRSVSEKTAQIINTQDLEKSHTKLOEBIPVLHBEQE 1290
 Db 1778 SESVPSAKSANPAVSKOPSSHDEINNYL-----QQID---QLKRIAGLEEEKQ 1823
 QY 1291 LLPNVKVSQETQETWNELELLELLETSQSTKSTTLARIEMERLRNLEKQESOEERKSLITKE 1350
 Db 1824 ---KNKFSQTL--NEKNYLLSQISTDG--ELKMLQOBEVTKMNLNQIIOELSRVTK-1876
 QY 1351 RDNKTTIKEALEVVKHDLKEHIRETLAKIQESOSKQESQKQSLNMKEDNETTKIVSEMEOFK 1410
 Db 1877 -----LKETAEBEKDLEERLNNQLAELNGSIGNYQDVTDAIQKNE-1919
 QY 1411 PKDSALLRIEIMGLSKRLQESDHDMKSVAKKDDIQLRQLEVLQSESDQKKEIKETVA 1470
 Db 1920 -----LESEMKNLKKCVSELEEEKQOLVKEKTV-----ESEIRKEVLEKIQ 1962
 QY 1471 KHLK--TEEBELKVAHCLCKEETINELRVNLSKEKETEISTIQKLEBAINDKLQNKIOEI 1528
 Db 1963 AQEPGNKSHAKLOELLKKEQEVQKQKQKQIRYQEKISALERTVKAL-----EF 2013

1529 YEKELQINIKOISEVQVNNELKQFKEHRKAKDSALQSTIE-----SKMLEL 1574
2014 VQTESQ---KOLEITKEN---LAQAVEHRRKKAQELASFKVLLDDTQSEARVLADNLKL 2067
1575 TNRLOQSEIEIQIMIKEKEE---MKRYQEAQIQRDQKENTKEIVAMKESQKEQYFLK 1632
2068 KVELQNKESVSKQKQKQEDLERLEQA---EELHKE-----KKNMQEKLDAIR 2115
1633 MTAVNETQKMCIEHL---KEQFETOKL---NLNIEETENIRLTQILHENLEEMRSYTKE- 1687
2116 REKVH-LEETIGETIQLANKKEVQQLQENLSDTQVLAFTKSSMSLODDRVIDEA 2174
1688 -----RDLRSVEETLKVREQ---LKENLRETTIRLEKQEBELKIVHMLKEH----- 1733
2175 KWERKFSDAIQSKEEIRUKEDNCVSKQLRQMSIH---MEELKI-NISRLHDKQI 2229
1734 -----QETIDKLRGIVSEKTEINISNMOKLEHNDALKAQDLKIQEBELRAHM 1781
2230 WESKAQTEVQLQKVCDTLQ---ENKELLSQLEETRHLYHSSQNELAKLESELKSLKD 2285
1782 HLKEQOBTIDK---LKGIVSEKTDKLSN-----MQKOLENSNAKLQEKIQBLKA 1827
2286 QLTDLNSLEKCKEQKQNLGIIIRQOEADIQNSFSEYQLETDIQAASRELTSLRHLSEINM 2345
1828 NEHQIITLKDQVNET-----QKKVSEMEQLKQIKQDQSL----- 1861
2346 KEQKIISLSSKEBAIQAIAELRQOHDKEIKELNLLSQEBEENIVLEENKKAVDKTN 2405
1862 -----TLKLEINENLAQELHENLEEMKVMKBRNL-----RRVEE---TLKLERDQKE 1910
2406 QLMETLTKIKENIQKAQDSFVKSSMSLQNDRIIVGDYQQLBEHLSHLEKQQL--- 2463
1911 SLOETKARDLIQOELKTARMLSKHEKETVDKLEKISEKTIQISDIQKOLDKSKDELQK 1970
2464 IQEAAANNNKLKEIRGLR---SHMDLSENNAKLDAELLOYREDLNQVITIKDSQK 2518
1971 KIOELQKELQLLRKVEDVNMSSHKKINEMQOLKQFEPNYLCKCEMDNFQTKKLHESLE 2030
2519 QLELVQ-----LQONKELENKYAKLEELKESEBAN-----EDLRRSFN 2557
2031 ETRIVAKERDELRRIKESLKMEDRQFIATREMIAARDQNHQVKEPKLLSQOQHLMES 2090
2558 ALQ---BEQDLSKEITESLKVSIQSL-----TRQVTAQOEEGTJ---GLYHAQK 2601
2091 LR-EKCSRKELL---KRYSEMDDHYEGLNRLSLDLKEIEFHRIMKKLY-----VL 2139
2602 VKEEVHRLSALFSSQKRAIELEELVQVKEAAKVGGEIE-DKLKELKHLHHDAGIM 2660
2140 SYVTKIKESQHECINKFEMDFIDEVQKQKELIKIQLQOQCDVPS--BELRDLKLNQNM 2197
2661 RNETETAER---VAELARDLVZ---MEQKLLMVTKENKGLTAQIQSGFRSMSSL---QNS 2712
2198 DLHIEBILKDFSEFPSTKTEPQVLSNRKEMTQPLEWINTRPDIKLNKIQKENDR 2257
2713 -----RDHANEELDELKRYDASL---KELAQLKEQGL-----LNERDA 2749
2258 ICQVNNFNNRIITAINESTEPERSATISKEMEQDLKS-----LKEKNEKLFKNYQT 2310
2750 LISETAFSMN-----ST--BENSLSHLEKLNQQLSKDBQLLHLSQLEDSYNQVQS 2799
2311 LKTSLASGAQVNPPTQNKPHVTSRATQLTTEKIRELE-----NSLH 2353
2800 FSKAMAS-----LQNERDLWNE-----LEKFRKSEBGKQSAAPSTSPAEVQSLK 2846
2354 EAKESAMHESKSIK-----MQKELEVNTDI---JAKLOAKVHESNKLEKTEITQVLOD 2406
2847 KAMSSLQNDRLKELKNLQOQYLQINQEIITELHPLKAQLOEYO---DKTK-AFQIMQE 2902
2407 KVALGAPYKEEIEDLKMVKIDLEKMKNAKEPEKISATKAVEYQKEVIRILREMLR 2466
2903 ELROENLSWOHELDQLRMEKSWEIHERRMEQYVLMASDKQQLSHLQNLIRELRS-- 2960
2467 RSQQAQDTSV-ISEHTDPQPSNKPPLTCGGSGGIVQNTKAL-----ILKSHIRLEKEIS 2519

2961 -SSQTPKLVQVQROASPETSASP---DGSQNLVYETELLRTQLNDSLKEIH-QKEIRIQ 3015
2520 KLIKQNEQLIKQKN-----ELLSNNOH---LSNEVTKWERTILKRAHKQVTC 2565
3016 QLNSPNSQLLEKNLTLSIQLCDTSQSLRENQOQHYGDLNLHCAVLEKQVQELQA-GFLNID 3074
2566 NSPKSPKVTGTASKKQIITPSOCKERNLQDPVPKESPKSCFFDSRSKSLPSPHPVRYFDN 2625
3075 VAPGAPOEKNQVHRK-----SDPELREPQQSFSEAOQQLCNTRQEV---N 3117
2626 SSLGLCPFV---QWAGAESVDS-----QPGFWHASSGKVPECKTO 2663
3118 ELRKLLEERDQVAAENALSVAEQIRLREHSEWDSRTPPIIGSCGTQ 3166

RESULT 4
US-10-332-089-2
; Sequence 2, Application US/10332089
; Publication No. US20040072175A1
; GENERAL INFORMATION:
; APPLICANT: Cytokinetics, Inc.
; APPLICANT: Beraud, Christophe
; APPLICANT: Sakowicz, Roman
; APPLICANT: Wood, Kenneth W.
; TITLE OF INVENTION: NOVEL MOTOR PROTEINS AND METHODS FOR THEIR USE
; FILE REFERENCE: 020552-001610US
; CURRENT APPLICATION NUMBER: US/10/332,089
; PRIOR FILING DATE: 2002-12-17
; PRIOR APPLICATION NUMBER: WO PCT/US01/16300
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: US 09/572,191
; PRIOR FILING DATE: 2000-05-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patent in ver. 2.1
; SEQ ID NO 2
; LENGTH: 1388
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Hskif15 predicted amino acid sequence.
US-10-332-089-2

Query Match 8.1%; Score 1077.5; DB 12; Length 1388;
Best Local Similarity 26.5%; Pred. No. 1.1e-45;
Matches 475; Conservative 280; Mismatches 521; Indels 519; Gaps 79;

QY 3 BEGAVAVCVRVRLPNSREESL-GETAQVYVKTQNNVIVQVDCS-----KSFNED 50
Db 23 EGDALVFVRIRPPAERSGSADGE-----QNLCLSVLSSTSLRLHSNPEPKTFTFD 73
QY 51 RVFHGNETKNVYEETAAPIIDSAIOGYNGTIFAYGQTASGKTYTMMG-----SEDL 103
Db 74 HVADVDTTQESVPATVAKSIVSECMGNGTIFAYGQTSGKTFMMGFSSESNFHNLR 133
QY 104 GVTPRAIHDF-----QKIKKFPDRFLLRVSYMEIYNETITDLCGTOQMKPLIREDV 158
Db 134 GVTPRFYFSLIDREKAKAGKSFCKCSFIIEVNEQIYDLL--DSASAGLYREHI 191
QY 159 NRVVYVADLTEEYVYVSEMALKWITKGEKSRHYGETMMQORSRSHITFIRMILESREKGE 218
Db 192 KGVFVVGAVQVVTSAEAYQVLSGGRNRRVASTSMNRSRSHAVFTITTESMEK-- 249
QY 219 PSNCEGSKVSHNLVDLAGSERAATGAAGVRLKEGCNINRSILFGQVVKKLSQGVG 278
Db 250 -SNEIVNIRTSLLNVDLAGSERQKDTHAEGMRKEAGNINRSLSCLGQVITALVD--VG 306
QY 279 G----FNYVDSKLTRILQNSLGNPKTRIICTIPVS--FDETLTALQFASAKYKNT 332
Db 307 NGKQHVYRDSKLTFLRLDSLGGNAKTAIIANVHPCGRCFGETLSTLNFQAKLKNK 366
QY 333 PYVNE-----VSTDEALLKRYKREIMDLKK-OLEEVSLETR-----AQAMEKDQALQL-- 380

Db 367 AVVNEDTQGNYSQQAQVRLKEQLAELASGQTPPEFLTRDKKTKNTMYEQEAMLEFFK 426
 Qy 381 -----BEKDLLOKQVNEKIENTRLMLVTSSTLQOELKAKKRRVTCGLGKINKMGKNY 436
 Db 427 KSEQEKKSLEIKV-----TQLEDLTLKKEKFI-----QSNKMIVKFR 463
 Qy 437 ADQENIPNITTKTKLSI-NLLREIDSVCSDFVFNITDITLSE-IEMNPATKLLNOE 494
 Db 464 EDQI-----IRLEKHKESRGGLPEEQDRLLSE---LRNEIQTLEQIEHPRVAKYAME 516
 Qy 495 NIESELSRADYDNLVDYQLARTEKEEMELKLEKNDLDEPFALEKTKTKDQEMQLIH 554
 Db 517 N-----HSLR-----PENRRRLLE-----PVKRAQEM----- 539
 Qy 555 EISNLNLVHREYVNOLENELSSKVLELREKEDQIKLQYIDSQKLENIKMDLSVSL 614
 Db 540 ---DAQTIKLEKAF-----SEISG-----MEKSD---KNQGF----- 567
 Qy 615 ESIEDPKOMQTFLDAETVALDAKRESAFLRSENLLEKEMK-ELATYKQEMENDIQLYQ 673
 Db 568 -----SPAQKEPCLFANT-----EKLKAQLQIQTELNNKQEYE 603
 Qy 674 SOLEAKKMOVDLEKELQSAFNEITKLTSLIDGKVPKDLNLCNLEGEKITDIQK-ELNKE 732
 Db 604 EFKELTEKROLESELSQ-----LOKANLNLE 631
 Qy 733 --VEENEALREEVILLSELKSLPSEVERLRKEIQDKSEELHIITSEKDKLFSEVVKESR 790
 Db 632 NLLEATKACKRQ-----EVSQNLK---IHAETIKLIITP-----TKAYQLHSR 671
 Qy 791 VQGLLEIGTKDKDLATQSNYKSTQEQFONFKTLMDPEQKYKMWLE---NERMNOEI 847
 Db 672 -----PVFKLSEPMGSGFLYT-----QNSILND-----ILNEPVPEMNEQAF 712
 Qy 848 VNLSEAKQFDSGLGAKFISLQEKNELOQMLESVIAEKEQLKTDLKENIEMTENO 1084
 Db 713 EALISELRVTQEQMSALQAKL---DEEHNKLCQOHVDKBBHSTQMOELFS----- 762
 Qy 908 REKTLITEKLOQTELEVVTLOKDKLQLOESLOIERDOLKSDIHD---TVANNIDTQEQ 965
 Db 763 -----SERIDWTKQOELLSQLNVLEKQLOET-QTKNDFLSEVHDRLVLSADKE--- 813
 Qy 966 LRNALESKHOQTINTLASKISEEVS-RNLHMEENTGETKQEFQKMGVIGDKQDLEAK 1024
 Db 814 ---JSSVKLEYSFKTNOKEFNKLSERHMHV-----QLQDLNRLNEKLESK 860
 Qy 1025 NTQTLTADYKNEIIEQQRKIFSLQEKNELOQMLESVIAEKEQLKTDLKENIEMTENO 1084
 Db 861 ---ACLQDSYDN-LQEIIMKF-----EIDQLSRNLQNFKENETLKSDDLNNLMEL-LEAE 909
 Qy 1085 BELRLILGDELKQOEIVAKQKHAIKKEGELSRTCDRLAEVEBEKLEKESQQLQEQOQLL 1144
 Db 910 KE---RNNKLSIQFE---EDKENSCK-----EILKVLAEVQEQKQETA 947
 Qy 1145 NVQEEMSMQKINIEINLKNELKNELTLEHMETERLELAQKLNENVEEVKSYTKERKV 1204
 Db 948 KCEQMAKQVK-----LE-----ESLLATEKV 969
 Qy 1205 LKELQKSPETBRDLHRLGYRIEIEATGLQTKBELKIAHLKHEQETIDELRSVSEKTAQ 1264
 Db 970 ISSLEKSRDSKVVADLMNQIQ-----ELRSSVCEKT-E 1003
 Qy 1265 LINT-QLEKSHTKL-----QREIPVLHBEQELLPNVKVSETOETMNELELITQEST 1316
 Db 1004 TIDLKQELKDQINCYNALSALVDRESRVLIKKQEV-----DILDLKETL-RRLILSE--- 1054
 Qy 1317 TKDSTTLARIEMERLRINEKQESQEBEIKSYTKERDNLKTIKEALEVHKDQKHEIRETL 1376
 Db 1055 -----DIEDRMLCEDLAHATEQLNMLTE-----ASKKSHGLQSAQEL 1093
 Qy 1377 AK-----IQESQSKQEQSLNMKEDNETTKYIVSEMEQPKPKDSALLRIEIMGLSKRLQE 1432
 Db 1094 TKKEALIQELOHK---LNQKKE-----EVEQKQNE-----YNFKMRQ 1127

Qy 1433 SHDEMKSVAEKDKDQLQRLQEVLOSQESDQKLENIKEIVAKHLET-EEELKVAHCLKEQBE 1491
 Db 1128 LEHYMDSAAED-----PQSPKTPPHFQTHLAKLLETQOEIEDGRASKTSLSH 1175
 Qy 1492 TINELRNLVSKETEISTIQKOLEAINDKLQNKQIEIYEKEEOLNIKOISEVQENVNBLK 1551
 Db 1176 LVTKLNDREVRNABILRMKEQLREM-ENLSQSLIEKNWLL-----QGQLDDIK 1226
 Qy 1552 QFKEHRKAKDSALQSIIESKMLE-LTNRLQESQEBIQIWKKEBEMKRYQEAL----- 1602
 Db 1227 ROKENSQNHDPNQLKNEQESIKERLAKS-KLVEEMLKMAADLEEVQSALYNKMECL 1285
 Qy 1603 ----QIREDQKLENTKEIVAKMKEQSOKEYOFLKMTAVNETQEKQC-EIEHLKEQFETQK 1657
 Db 1286 RMTDEVERTQTLSE-----KAFQEKQELRSKLEEMYERERTSOEMEMLRKQ----- 1332
 Qy 1658 LNLNIEETENIRLT--OILHENLEEMESVTKERDDLSRVEETLKVRRDQKLENLR 1710
 Db 1333 --VECLAEBENGKLVGHQNLHQKIQYVVRLLKENVRVLAEBETEKLRAENVFLKEKR 1385
 RESULT 5
 US-10-146-473-82
 ; Sequence 82, Application US/10146473
 ; Publication No. US2003010888A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Scanlan, Matthew
 ; APPLICANT: Gout, Ivan
 ; APPLICANT: Stockert, Elisabeth
 ; APPLICANT: Gure, Ali
 ; APPLICANT: Chen, Yao-Tseng
 ; APPLICANT: Old, Lloyed
 ; TITLE OF INVENTION: Breast Cancer Antigens
 ; FILE REFERENCE: L00461/70130(JRV)
 ; CURRENT APPLICATION NUMBER: US/10/146,473
 ; CURRENT FILING DATE: 2002-05-15
 ; PRIOR APPLICATION NUMBER: US 60/291,150
 ; PRIOR FILING DATE: 2001-05-15
 ; NUMBER OF SEQ ID NOS: 82
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 82
 ; LENGTH: 1388
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-146-473-82
 Query Match 8.1%; Score 1077.5; DB 14; Length 1388;
 Best Local Similarity 26.5%; Pred. No. 1.1e-45;
 Matches 475; Conservative 280; Mismatches 521; Indels 519; Gaps 79;
 Qy 3 BEGAVAVCVVRPLNRSRESL-GETAQVYWKTDNNVIYQVDS-----QNLCLSVLSSTLRHLSNPEPKITFTD 73
 Db 23 EGDALKVFRIRPPAERSGSADGE-----QNLCLSVLSSTLRHLSNPEPKITFTD 73
 Qy 51 RVFHGNETTKNYYEIAPIIDLSALQNGNITPAYGQTASGKTYTMMG-----SEDLH 103
 Db 74 HVADVDTTQESVFATVAKSIVESCMGNGTIFAYGQTGSGKFTTMMGSPESDNFSHNR 133
 Qy 104 GVIPRAHDIF-----OKIKKFPDRELLRYSYMEIYNETITDLCGTQKMKPLIIRDV 158
 Db 134 GVIPRSEYLSLDRKEKAGAGKFLCKCSFIEIYNEQIYDL--DSASAGLYLREHI 191
 Qy 159 NRVVYADLTBEVVYVSEMALKWITKGEKSHYGETKMNQSSRSHTIFRMLISREKGE 218
 Db 192 KKGVVVGAVEQVTSAAAYQVLSGGWRNRVASTMNRSSSHAVFTIISMEK-- 249
 Qy 219 PSNCEGSVKSHMLNVLDAESRAAQTGAAGVRLKEGCNINRSILFQILQVILKLSLSDGVG 278
 Db 250 -SNEIVNIRTSLLNVLDAESRQKTHABGMRLKEAGNINRSILCLGQVILTALVD--VG 306
 Qy 279 G-----EINYRDSKLTRLQNSLGNPKTRICITIPVS--FDETLTALQFASHTAKYMKNT 332

```

307 NGQHVYCYRDSKLTILLRSLGNAKTAIIANVHFGSRFCGETLTLNFAQRAKLKNN 366
333 PYNE-----VSTDEALLKRYKEIMDLKK-OLEVSLSTR-----AQAEMKQOLAQL-- 380
367 AVNEDTOGNVSOLOAEVRLKQELAEASGOTPPESFLTRDKKKNYMEYFQEMLFFK 426
381 ----BEKDLLOKQNEKIENITMLVTSSSLTLQELKAARKRVRWCLGKINMKNNSNY 436
427 KSPQEKSLTEK-----TOLEDTLKKEFI-----QSNKMVKPR 463
437 ADOFNIPNTITTKHLSI-NLAREIDSVCSRSDFVSNLTDLTSE-IENWPAKLLNQE 494
464 EDQI----IRLEKLHESRGFLPEEQDRLSE--LRNEIQTILREQIEHHPVAKYAME 516
495 NIESELNSLRADYNVLVDYQELTEKEEMELKLKEDNDLDEFALERTKTKQEQMLIH 554
517 N-----HSLR-----BENRRLLLE-----PVKRAQEM----- 539
555 EISNLKLVHREVVYNQDLENELSSKVELLREKEDQIKKQLEYIDSOKLENIKMDLSYL 614
540 ---DAQTIAKLEKAF-----SEISG-----NEKSD--KNQOGF----- 567
615 ESIEDPKOMQTLDAETVALDAKRESAPLSENLELKERKK-ELATTYKOMENDIQLYQ 673
568 ---SPKAQKEPCLFANT-----EKLKAQLLQIOTELNNSKQOYE 603
674 SOLRAKKQOVDELEKLOSAFNEITKLTSLIDGKVPKDLLCNLEGGKITDLOK-ELNKE 732
604 BFKELTRAROLESELOS-----LOKANLNLE 631
733 --VENEALREVVILLSELKSLPSSEVERLKEIOPKSBELHIITSEKDLFSVVHKESR 790
632 NLLEATKACKRQ-----EVSQNLK--IHAETKLIITP-----TKAYQLHSR 671
791 VQGLLEIEGKTDDIATTQSNKYSTDQEPONFKTLHMFQEQKVVLEZ--NERMNQEI 847
672 -----PVPKLSPEMGSGSLYT-----QNSILDND-----ILNEPVPEMNEQAF 712
848 VNLSEAKFDSGLAKTELSYKTOELOEKTREYQERLNEQKLEQLENRDSPLQTV 907
713 EAISEBELRTVQPMGALQAKL-----DEEHKMLKQOQVHDKLBHHSITQOELFS----- 762
908 REKTLITEKLOTEEVKTLTKQDKLQOLBSLQTERDOLKSDIHD--TVNMNIDTQBO 965
763 -----SERIDWKQOELLSQLNVLEKQLOET-OTKNDPLKSEVHDLRWLHLSADKE-- 813
966 LRNALESKQOHTINTLKSKEEVS-RNLHMEENTGETKDFEQKMGVIGDKQOLEAK 1024
814 ----LSSVKLEYSSFKTNOEKEFNKLSERHMHV-----QLQDLNLRLENEKLESK 860
1025 NTQTLTADVKDNEIIEQOKIFSLIOKNELOQMLSEVIAEKEPOLKTDLKENTEMTIENQ 1084
861 ----ACLOQSYDN--LOEIMKF-----EIDQLSNQLNFKKENETLKSDDLNNLMEL-LEAE 909
1085 EELRLLGDELKQOETVAOENKHAIEKESLRTCDRLAEVEBKIKESQOLQEQOQLL 1144
910 KE--RNNKLSLOFE--EDKENSCK-----EILKVLAEVROKOKETA 947
1145 NVQEMSEMOKKINEIENLNKELNKNKELTLEHMETELELAQKLNENVEEVSITKERV 1204
948 KCEQMAKVOK-----LE-----ESLATEKV 969
1205 LKELOKSFEDRHLRGVIREIATGLQTKELKIAHILKEHOETTIDELRRSVSEKTAQ 1264
970 ISSLEKSRDSKVVADLMNQIQ-----ELRTSVCEKT-E 1003
1265 IINT--QOLKESHTKL-----QBEIPVLHBEQBLPNVKKVSTQETMNELELLTQST 1316
1004 TIDLKQELKDINCYNKNSALVDEESVLIKKQEV-----DILDKETL-RLRLISE-- 1054
1317 TKDSTTLARIEMERLRNEKTOEQOEBEIKSTKERDNLTKTKEALEVVKHDLQKHEIRETL 1376
1055 -----DIERDMLCEDIAHAHQNLNTE-----ASKHSGLLQAQEL 1093

```

RESULT 6

```

US-10-173-999-32
; Sequence 32, Application US/10173999
; Publication No. US20040005563A1
; GENERAL INFORMATION:
; APPLICANT: Mack, David H.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Ovarian Cancer, Compositions
; TITLE OF INVENTION: and Methods of Screening for Modulators of Ovarian
; TITLE OF INVENTION: Cancer
; FILE REFERENCE: 018501-002420US
; CURRENT APPLICATION NUMBER: US/10/173,999
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: US 60/299,234
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: US 60/315,287
; PRIOR FILING DATE: 2001-08-27
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/372,246
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 163
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 32
; LENGTH: 1388
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-173-999-32

```

```

Query Match      8.1%; Score 1077.5; DB 15; Length 1388;
Best Local Similarity 26.5%; Pred. No. 1,le-45;
Matches 475; Conservative 280; Mismatches 521; Indels 519; Gaps 79;

QY      3  EGCAVAVVRVRPLNRSSESL-GETAQVYWKTDNNVYQVDSG-----KSFNFD 50
DB      23  EGDAIKVVRIRPPAERSGSGADGE-----QNLCLSVLSSTLSRLHSNPPKPTFFD 73
QY      51  RVFHGNETTKNVVEETAAPIIDSAIOYNGTIFAYGOTASGKYTWG-----SBDHL 103
DB      74  HVADVDVTTQBSVFATVAKSIVSCMGYNGTIFAYGOTGSGKFTWMPGSDNFNSHLR 133
QY      104  GVTPRAIHDIIF-----QIKKFPDRELLRVSVMEIYNTITDLCGQTKMKPLIREDV 158
DB      134  GVIPRSEYFLSLIDREKEKAGAKSFLCKSFIEIYNEQIYDLL--DSASAGLYLREHI 191
QY      159  NRVYVADLTVEVYVYSEMALKWITTKGKSRHYGCTQNNQSSRSHTIFRMILESREKGE 218

```


23 EGDAIKVVRIRPPAERSGADGE-----QNLCLSVLSTSLRLHNSPKNPFTTFD 73
 51 RVPHGNETTNYVEETAAPIIDSAIOYNGTIFAYGOTASGKTYTWG-----SEDLH 103
 74 HVADVDVTOESVPATVAKSIVSCMGYNGTIFAYGOTGSGKTYTWGSESNFPHNL 133
 104 GVTPRAHDF-----QKIKFPDRFLLRVSYMEIYNETITDLCGOTQMKPLIREDV 158
 134 GVTPRFEYLFSLIDREKAGAGKAGKFLCKSCFIEIYNEQIYDLL--DSASAGLYLREHI 191
 159 NRVVYVADLPEVVYVYSEMALKWITTKGKSRHYGVTGQNRSSRSHITFIMILESEKGE 218
 192 KKGFFVVGAVEQVTSAAEYQVLSGCMWRNRVASTSMNPRESSRAVFTITTESMBK-- 249
 219 PSCGEGVKVSHNLVLDLAGSRAAQTGAAGVRLKEGCNINRSILFGLGVIKKLDQGVG 278
 250 -SNEIVNIRTSLLNVLVLAGSEROKDTHAEGMRUKEAGNINRSLSCLGVITALVD--VG 306
 279 G-----FNYRDSKLTILONSLGNPKTRIICTITPVSS--FDETLTALQFASAKYMWNT 332
 307 NGQRHVICYRDSKLTFLRDSLGNKTAIIANVHPGSRGCFGETLSTLNFPAQRAKLINNK 366
 333 PYNE-----VSTDALKRYKREIMDLKK-OLEEVSLETR-----AQAMEKDOLAQL-- 380
 367 AVNVEDTQGNVSQLAQVRLKEQALASQTPPESTFLTRDKKKTNYMEYFQEAULFFK 426
 381 ----EEDOLLQVONEKIENITRMLVTSSSTLQOELKAKRKRVTWCLGKINKMKNSNY 436
 427 KSEQEKSLIEKV-----TOLEDLTLKKEKFI-----OSNMIVKFR 463
 437 ADOFNIPNITTKHLSI-NLLREIDSVCSSESVFNTITDLTSE-IEWNPATKLNLQR 494
 464 EDQI-----IRLEKLHESRGFLPEEQDLISE--LNEIQTLEQIEHPRVAKYAME 516
 495 NIESELNSTRADYDNLVDYQLRTEKEMEELKKEKNDLDEFEALERKTKQOEMQLIH 554
 517 N-----HSLR-----EENRRLLE-----PVKRAQEM----- 539
 555 EISNLKNLVKREYVNOLENELSKVLELLREKEDQIKKQVYDSQKLENIKMDLSYSL 614
 540 ---DAQTIKLEKAF-----SEISG-----MEKSD--KNQOFP----- 567
 615 ESIEDPKQKOTFLDAETVALDAKRESAFLRSENLELKEKMK-ELATTYKQWENDIQLVQ 673
 568 ----SPKAQKEPCLFANT-----EKLKAQLLIQIOTELNNSKQEYE 603
 674 SOLEAKKKQVQVLEKELQSAFNEITKLTSLIDGKVPKOLLNLELECKITDLOK-ELNKE 732
 604 EFKELTRKQLESELOQ-----LOKANINLE 631
 733 --VEENEALREEVILLSELKSLPSEVERLRKEIQDKSEBILHITSEKDLFSEVHVHESR 790
 632 NLLEATKACKRQ-----EVSQNLK--IHAETKLIITP-----TKAYOLHSR 671
 791 VQGLLEIEGKTODDLATTOGNYKSTQDFQNFKTHMDPEQKYKMWLEE--NERNWOEI 847
 672 ----PVKLSPEMSGFSGLYT-----QNSSILND-----ILNEPVPPMNEQAF 712
 848 VNLSEKQAFDSSLGALKTBLSVKTQELQKTEVQERLNEMLQKLEENRDSPLQTVSE 907
 713 EATSEELRTVOEQMSALQAKL-----DBEEHKNLKQOHVDKLBHHSTOMQELFS----- 762
 908 REKTLITEKLOOTLEEKVTITQEKDDLQOLQESLIQTERDQKSDIHD--TVNNNIDITQEQ 965
 763 ----SERIDWTQKQBELLSQLNVLKQLOET-OTKNDFLKSEVHDLRVVLHSADK-- 813
 966 LRNALSLKQHOETINTLTKSISEVS-RNLHMEENTGETKDFQOQWGWIDKQDLK 1024
 814 ----LSSVKLEYSFKTKQEFNKLSERHMHV-----OLQDLNLRLENEKLESK 860
 1025 NTQTLTADVKNDNEIIEQORKIFSLIQKNEVLOQWLESVIAEKEKQKTDLENIENTENQ 1084
 861 ---ACLOQSDYN--LOEIMKF-----BIDQLSRNLQNFKKENETLKSOLDNNLMEL-LEAE 909

1085 BELRLGDLKQOEIVAOEKNHAIKKEGELSRTCDRLAEVEBKLEKESQOLQEQOOLL 1144
 910 KE---RNNKLSLOFE---EDKENSX-----EILKVLAVRQEKOKETA 947
 1145 NVOEEMSEMQKINEIENLNKLNKELTLEHMETERLELAQKLNENYEVKSIYTERKV 1204
 948 KCEQMAKVQK-----LE-----ESLLATEKV 969
 1205 LKELQKSFETERDHLRGYIREIATGLQYKEELKIAHILHKEHOETIDELRRSVSEKTAQ 1264
 970 ISSLESRSDSKKVVADLMNQIQ-----EELTSVCEKT-E 1003
 1265 IINT--QDLKSHTKL-----OEETPVLHEBOELLPNVKKVSETOFTMNELELLTPOST 1316
 1004 TIDTLQOELKDINCKNSALVDRSESVLKKQEV-----DILDKETT-LRLILSE--- 1054
 1317 TKDSTTLTARIMERLRLNEXFQSEBEIKSLTVERDNLKITKEALEVKKHOLKEHIRETL 1376
 1055 -----DIERDMLCEDLAHATEQLNMLTE-----ASKKHSGLQSAQOEL 1093
 1377 AK-----IOESQSQOBSQINMKKDNETTKIVSEMOQFKPKDSALLRIEIMLGLSKRLOE 1432
 1094 TKKEALIQELQHK-----LNQKKE-----EVEQKNE-----YNFKMRQ 1127
 1433 SHDEMKSVAKEKDQLRLQELVLOQESDQKLENIKEIIVAKHLET--EELKVHACCLKQOBE 1491
 1128 LEHYMSAAED-----PQSPKTPPHFQTHLAKLETQOETIEDGRASKTSLEH 1175
 1492 TINELRVNLSEKETEISTIQOLBAINDKQNKQIOTIYKEEEOQLNIKOISEVOENNELK 1551
 1176 LVTKLNEEDREVKNAILRMKEQLREM-ENLRLESQOLIEKNWLL-----OQQLDDIK 1226
 1552 QFKHRKAKDSALQIESKMLE-ITNRLQESQOBEIQTIMIKEEMKRVQOEL-- 1602
 1227 ROKENSQDHPDNOQLKNEQESIKERLAKS-KIVEEMLKWADLEEVQSALYNKEMECL 1285
 1603 ----QIERDOLKENTKEIVAKMKBSQEKYQFLKMTAVNETQKMC-BIEHKEQFETQK 1657
 1286 RMTDSEVETOTLES-----KAFQEKQLSKLEEMVEERERTSQEMEMLRKQ----- 1332
 1658 LNLNIETENIRLT--QILHENLEEMRSVTKERDDLRSDRSVEETLKVVERDOLKENLR 1710
 1333 --VECLAEEGKLVGHQNLHOKIQYVVRKKNVRLABETEXKLRAENVELKEKR 1385

RESULT 8

US-10-171-311-4
 ; Sequence 4, Application US/10171311
 ; Publication No. US20030087270A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Schlegel, Robert
 ; APPLICANT: Chen, Yan
 ; APPLICANT: Zhao, Xumei
 ; APPLICANT: Monahan, John
 ; APPLICANT: Kamatkar, Shubhangi
 ; APPLICANT: Glatt, Karen
 ; APPLICANT: Gannavarapu, Manjula
 ; APPLICANT: Hoerish, Sebastian
 ; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
 ; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
 ; TITLE OF INVENTION: OF CERVICAL CANCER
 ; FILE REFERENCE: MRI-035
 ; CURRENT APPLICATION NUMBER: US/10/171,311
 ; CURRENT FILING DATE: 2002-06-12
 ; PRIOR APPLICATION NUMBER: US 60/298,159
 ; PRIOR FILING DATE: 2001-06-13
 ; PRIOR APPLICATION NUMBER: US 60/298,155
 ; PRIOR FILING DATE: 2001-06-13
 ; PRIOR APPLICATION NUMBER: US 60/335,936
 ; PRIOR FILING DATE: 2001-11-14
 ; NUMBER OF SEQ ID NOS: 238
 ; SOFTWARE: Fast-SEQ for Windows Version 4.0

```

; SEQ ID NO 4
; LENGTH: 3899
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-4

Query Match      8.0%, Score 1064; DB 14; Length 3899;
Best Local Similarity 19.7%; Pred. No. 1.9e-44;
Matches 672; Conservative 618; Mismatches 1107; Indels 1010; Gaps 150;

QY 26 TAQVYKTDNNVIYGVYDYSKSFNDFRPHGN-----ETTKVYBEIAPIIDSAIQYNGT 81
DQ 295 TWQISFLQEKIKVYEMQDK-----KVNSKEBEIQKETIIBELNATIIIE----- 341
QY 82 IFAYGOTASGKTYMMG-----SEDLGLGVI PRAIHDFQIKTKPPDREFFLLRVSTMEYN- 136
DQ 342 -----EKKTELEKDKLTADKLLGELQEQVQKQEIKNK-----LEITNS 383
QY 137 -----ETITDLCGTQKMKPLIITREDVNRVYVADLFEVYVYTTSEMALKWITTKGS 188
DQ 384 KOKERSSBEIKQMGTVBELQ-----KRNHKSQFETDIVORMEQETQKLEQL-RAELD 438
QY 189 RHYGE--TKMQRSSSHITFRMILSRKGEPSNC-----EGSVKVSHLNLDVL 236
DQ 439 EMYGQIVQMKQELIROHMAQMEEMKTRHGMENALRSYNTVNEEDQIKLMVAINEL 498
QY 237 -----AGSERAAQTGAAGVRLKEGNCINRSLPILQVVKLSLSDG-----OVGFFINYRD 285
DQ 499 NIKLQDTSQKKEKLEELGLILEKCALQKQ---LEDLVEELSFSEQIQARQTTAQQE 555
QY 286 SKLTRILQNSLGNPKTRIICTITP-----VSPDETALTALQFASTAKYMKNTPYNEVST 340
DQ 556 SKL-----NEAHKSLSTVEDLKABIVSASESRKELE-----LKHEAEVINYKI 598
QY 341 DBALAKRYRKEIMDLKQLEEVSLR---TRAQAMEKDQLAQLLEKDLQKQNEKIENL 397
DQ 599 KLEMLEKKNVLDMAESQBALERLTQLFSEHEELSKLKDLEIEHRINIEK----- 654
QY 398 TRMLVTSSSLTIQOELKAKRRVTCWLGKINMKNSYADQFNPTNTITTKTK--LGI 455
DQ 655 -----LKNLGIHYKQID--GLQNEMSQKIETMQFE-KDNLTITKQNLILIEI 699
QY 456 NLLREIDES-VCSQSDVFNSTLDTL-SEIENPDKLQNEINRSELNSLRADYDNLVLD 513
DQ 700 SKLDLQOOSLVNSKSEEMTLQINELQKETE-----ILQOE--EKEKGTLEQEVQEL-- 748
QY 514 YEQLRTEKEEMELKLEKNDLDE-----536
DQ 749 --QLKTELLEKQKKE-NDLQEKFAQLEAENSILKDEKKTLEDMLKIHTPVSQERLIF 805
QY 537 PEALERTK-----KQEMQLIHEISNLKNLV-----KHREVNQDLE 574
DQ 806 LDSIKSKSDSWEKEIEILTEENEDLQOQIQLENEIEKQENTSPAKKPFVNYQELQ 865
QY 575 NELSKVELLEKEDQIKKQBYIDSQKLENKMDLSYLSIESIEDPK-OMKQTLFD----- 629
DQ 866 EYACLLKVKODLEDS-KNQOEYKSKLKALNEEL--HLQRIKNTTVKMSVFEDEKT 922
QY 630 --AETVAL-----DAKESAFIRSENLEKMKELATTYKQENDIQLOSQLEA 678
DQ 923 FVAETLENGEVKEDTTLEMKLEVTTRKLELSQRLSDLSLQKQKGEISFLNEEYKS 982
QY 679 KK--KMQVDLE-KELQSAFNE-----ITKLTSLIDGKVPKDLLCNLEKGTIDLQ 726
DQ 983 LKQEKQVSLRCRELEIINHNRAENVQSCDTQVSSLLDGVV---TWTSRGAESVSQVN 1039
QY 727 KELNKE-----VEE-----NEALREEVILLSELKSL-----PSEBRLRKEI 763
DQ 1040 KSGFEESKIMVEDKVSFENMTVGVESQEQILDLPLSPVTKESSLRATOPSENDKQLQEL 1099
QY 764 QD-KSEE-----LHI-----ITSEKDK-----LPSEVTH-KESRVQ 792
DQ 1100 NVLAKSEQNDLEQAEORICLSLVSTHVQOVREYMEKEDKALCSLKEELIFAQEBKIK 1159

```

1614 KEIVAKESQOEKEYOFLKMTAVNETQKMCIEHLKEQFETOK-----LNLENIENTENIR 1669
2195 KEIT-----NLEBOLEQFRE-----ELENKNEVOQLHMQLEIQKKESTTRLOLEQEN-- 2243
1670 LTOILHENLEEMASVTKERDILASVBETL-----KVERDOLKENL----- 1709
2244 --KLPDDMEKGLAIKESDAMSTODQHVLFGRFAQIQOEKEVIDQLNEQVTKLOOQLK 2301
1710 -----RETTIDLEKQOEELKIVHML--KEHQETIDKLRGIVSEKTNISNM-OK 1756
2302 ITTNDKVIENKELIRDETQIECLMSDQECVRNREEIEQLNEVIEKLQOELANIGQK 2361
1757 -----DLEHSDALKAQDLKIOBELRIAH----- 1780
2362 TSMNAHLSBEASLSKHLQDQVIAKLALEQQOVETANEEMTPMKVNLKTFNFKMNQLTQE 2421
1781 --MHLKEQOETIDKLRGIVSEKTD--KLSNMOKDLENSNAKIOELKANEHQ----- 1831
2422 LFSIKRERESVEKIQSPENSUNVAIDHLSKORPELEWLTE--DALKSLENQTPKSF 2479
1832 -----LITLKOVNETQKVS-----EMEQLKKQIKDOOS-----LTLKLE 1867
2480 ENGKSIINLETRILLOLESTVSAKDLELTQYKIQKDMQBGQOPETEMLOKQVNLQKIV 2539
1868 IENLNLAEHLHENLEEMKSVNK-----ERNLNRVRETLKLERDOLKESLOET 1915
2540 BEKVAALVSIQIEAEOYAKFQDNQTTISSEPERTNIQNLQ--LREDELGSDISAL 2596
1916 KARDLEIQQ---ELKTARMLSEKHEKTVDKLREKISEKTIQISDIQDLDKSKDELQK 1972
2597 TLRISELESQVEMHTSLILEKEQVEIAEK--NVLEKEKKLLOKLLEGNEKKORKE 2653
1973 QELQKKEQLLQVRVEDVNMHKK---INEMOLKQKQEPNYLCKEMDNFOLTKLHESL 2029
2654 KRSQPQVVLKTTTELFSNEBSGFFNEALRAE-----SVATKAEIASY--KEKAEL 2707
2030 EEIRIVAKE-----RDELRIKESLK--MERDQFIATLREMIARDRONHVKP 2075
2708 QE--ELLVKTNTSLQKLSQVRDHLAEKELKISLEKETEVEQESKACWFEPPIKL 2766
2076 EKRLIS--DGOOHLMES-----LREKES--RIKELLKRYSEMDDHYECL 2115
2767 SKSIASQTDGTLKISSNOTPOLVKNAGIQINLQSECSSEVTEIISQFTEKIEKMQEL 2826
2116 NRLS--LDLE--KEIEPHRMKUKVLYSVYTKIEB-----QHECINKFEMDFIDEV-- 2165
2827 HAAEILDMESRHISSETLTKREHYA--VOLKECGLTKAVIQLRSKEGSSIFELAHS 2884
2166 ---KOKELL-----IKIQLQDCDVPES--RELRDLKINQNDLHIEILKDFPS 2209
2885 DAYQREICSSDSGSDWGGIYLTH--SQGFDIASGRGESESATDSPPKIKGLLR--- 2940
2210 ESEFPSIKTEFOVLG-----NRKEMTOFLEWLNTR-----FDIEKLKNGLOK-EN 2255
2941 ----AVHNEGQVLSLTPSYSDGSDHDSIQVSEFWELEKAYINTISSLKLITRMQL 2995
2256 DRICQVNN-----FFNNRIIATMNESTE----- 2278
2996 QREAEFYDSSQSHEFSWGRGELLALQVFLERSVLLAAPTETALTGTDTDAVGLNC 3055
2279 FEERSATISKWEPQDLKSLKEKNEK--LFKNYQTLKTSLSAGAQVNPPTQDNKP----- 2331
3056 LEQRIQEOGVYQAAMECLQADRRSLSEIQAALHAQMNGRKITLREQESKPSOELLE 3115
2332 -----HVTSRAT-----OLITTEK--IRELENSLHEAK-----ESAMH 2361
3116 YNIQOQSOMLEMQVLESSLMKORATELOELSSSEKVMVAELKSELACTKLETTIKAOH 3175
2362 KESKIKMQK-----ELEVTNDILAKLQAKVHESNCKLEKTIETIQVLDQKVALGAK 2413
3176 KHLKELEAFRLVVKDQTDVHLLNDTLASEQKSLRELQWALEKEKA-----KLGRSEE 3228
2414 PYKEEIEDLKMVLKVIDLEKMKNAKEFEKISATKATVEYQKEVIRLLRENLRSSQAQD 2473

3229 RKEELEDLKFSL-----ESQQRNLQ-----LNLLLEQOKLLNESQOKIESQRMLYD 3277
2474 TVISISHTDPOPSNKLTCGGSGIVONTKALILKSEHILKLEKEISKLKQONE----- 2526
3278 AOLSEE---QGRNLEL-----QVLLSEKVRIREMSSTLDRERELHQAQLOS 3320
2527 -----OLIKOKNELLSNNOHLSNEVKTWKERTLK-----REAHK 2560
3321 SDGTGOSRRPPLSPEDILKLOLEEKHSHRIVELLNETEKYKLDLSQTRQOMEKQVHR 3380
2561 QVTCENSPKSPKVTGTASKKQITPPOCKERNLQDPVPKESPKSCFPDSSKSLSPSPHPV 2620
3381 KTL-----QTEQANTEGQKK--MHELQSKVEDLQORLEEKQVQVYKLDLEGQRLQ----- 3429
2621 RYFDNSLGLCPVQVQAGAESVD-----SOPGPHASSGK 2655
3430 -----GIMQBFQXQLEEREBKRESRRILYQNLNEPTTWSLTSR 3468

RESULT 9
US-10-171-311-8
; Sequence 8, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Ganuvarapu, Manjula
; APPLICANT: Hoersb, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 3917
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-8

Query Match 8.0%; Score 1064; DB 14; Length 3917;
Best Local Similarity 19.7%; Pred. No. 1.9e-44;
Matches 672; Conservative 618; Mismatches 1107; Indels 1010; Gaps 150;

QY 26 TAQVYKTDNNVYQVDSKSFNFRVFNHGN-----FTTKNVYEEIAAPIDSALQYNGT 81
DB 295 TWQISFLQEKIKVYEMEQDK-----KVNSNKEEIOEKETIIEELNTKILIE----- 341
QY 82 IFAYGQTASGKTYTWMG-----SEDLGLVLPRAIHDFQIKKFPDPREFLLRVSYMBIYN- 136
DB 342 -----EKKTUELKDXLTADKLLGELQEQIVQKNQEIKNMK-----LELTNS 383
QY 137 -----ETITDLCGTQMKPLIIREDDVNRNRYVADLTETEVVYVTSMAKWKITKXKS 188
DB 384 KQKERSSEIEKQLMGTVELQ-----KRNHKSQFETDIVORMEQETQKLEQL-RAELD 438
QY 189 RHYGE--TKNQBSRSSHIFRMLLESREKGEPSNC-----EGSVKVSHLNLVDL 236
DB 439 EMTYGOQIVQKQELIRQHMARQMEEMKTRHKGEMENALRSYNTVNEDQIKLMVAINEL 498

Db 806 LSIKSKSDSVWEKTEIIEENEDLKQOCIQNLNBEIEKORNTFFSAEKNFVNYQELQ 865
 Qy 575 NEUSSKVELLEREDQIKLQEIYDSQKLENKMDLSVLSIESIEDPK-QMKOTLFD--- 629
 Db 866 EYVACLKLVDDLEDS-KNQELVEYKSKLXALNEEL--HLQINPTTVKMKSSVFDDEKT 922
 Qy 630 --AETVAL-----DAKRESAFLRSENLKELKMKELATTYKQEMENDIQLYQSOLEA 678
 Db 923 FVAETLEMGVVEBKDTTLMELKLEVTREKLELSQRLSDLSQLKQKHGEISFLNEBVS 982
 Qy 679 KK--KMQVDLE--KELQSAFNE-----ITKLTSLIDGKVPKDLCLNLELGKITDLQ 726
 Db 983 LKQEKQVSLRCHLELIHNNRAENVQSCDTQVSSLLDGW---TMTSRGAESVSKVN 1039
 Qy 727 KEIANKE-----VEE-----NEALREEVILLSKLSL-----PSEVERLKEI 763
 Db 1040 KSRGEEKIMVEDKVSFENMTVGEESKQOEIILDHLPSTVTKESSLRATQFSEMDKQKEL 1099
 Qy 764 QD-KSEE-----LHI-----ITSEKDK-----LFSVVH-KESRVQ 792
 Db 1100 NVLKSEQNDLRLOMEAQRICLSIVYSTHVQVREYMEKDKKALCSLKEELIFAQEEKIK 1159
 Qy 793 GL-----LE-----EIGTKDDLATQSNYKST----- 815
 Db 1160 ELQKHQLELQTMKTQETDGEKPLHLILGKQKAVSEBYSFLOTLCSVLGEYTPALK 1219
 Qy 816 -----DQEPQNFKTLHMDPEQKYKVMLEB-NERMNOEIV----- 848
 Db 1220 CEVNAEDKENSQDYISENEDPELODYRYEQDQENWHILNKVTEYNKLVLOTRLSK 1279
 Qy 849 -----NLSKEAQKP-----DSSLGAL-----KTLSYK 871
 Db 1280 INGQOTDGMKLBFGREENLPKEETEFLSIHSQMTNLEDIDVNHKSLSLQDLEKTKLEQ 1339
 Qy 872 TQLOEKTREVERLNEMLQKLENRDPSPLQTVREKTLITEKLOQTLIEVKTUTQEK 931
 Db 1340 VOELSELSSLOQQLKETEQ-----NYEARHCLQKRLQAVSESTVPPSPVDSVITE 1393
 Qy 932 DDLKQLESQIATQERDQKSDIHDVTNMNTDTQBLRNALLESKHOETINTLSKI----- 987
 Db 1394 SDAQRTMYPGSCVKNIDGTIFSGFGEYKETEINIVKLE--KOYQOELEEEVAKVIVSM 1451
 Qy 988 -----SESVNLMHEWNTGTQDEF-----QQRWVGIDKQDLRAKNTQTL-TADVK-- 1034
 Db 1452 STAFACQOTELSRISGKENTASSKQAHAVCQOEQHYFENMKLSQDQIGQTFETVQVKEK 1511
 Qy 1035 -----DNEILBQKIFSLIOEKNELQMLSVIAEKEQL-----KTDLENTEM 1079
 Db 1512 BEFKPLSELGHBGKEILLNSDPHDIPESKDCVLTISEMFSKDTFTVRSIHDEISV 1571
 Qy 1080 -TIENQOEELRLGDELK-KQOIVAQEKNAHAIKKEGEL-----SRTCDRLAEVEEKLEK 1132
 Db 1572 SSMDASRQMLNNEEQLEDWRLVQYQEH--QOATELLRQAHMQRQERQEQOLQEE 1629
 Qy 1133 SQOLOEQOQLNVQEE--MSEMOK-KINEIENLKN-----ELKN----- 1169
 Db 1630 IKRLNRQLAQRSSIDNENLVSERERVLLEELKQLSLAGREKLCCELNRSSTQNGN 1689
 Qy 1170 -----KELTLEHETERL--ELAQKL--NENYEVKSITKERVKLQKQSF-----ETE 1215
 Db 1690 ENQGEVEEQTFKEKELDRPDPVPPFELSERNYALQKANNRLKTLLEVVKTAAVEETI 1749
 Qy 1216 RDHLRGYI-----RETEATGL--QTEELKIAHILKHEQETIDELRRSVSEKTAQIINTQ 1269
 Db 1750 GRHVILDRSKSQSSASLWRSAAEASVSKVHEETRVUDE--SIPSYSGDMPRN 1806
 Qy 1270 DLE-----KSHTKLOEBE-----PVJHEEQELLPNVKVSQTQMTNE--LELLITEQ 1314
 Db 1807 DINMWSKVTEBGTSLQSLVRSGFAGTEIDPENEELMLNIS--SFLQAAVEKLELAISET 1864
 Qy 1315 STTKDSTTLARIEMERLRLNEKFOESQOEIISLTKERDNLKTIKALEVXHDQKHEIRE 1374
 Db 1865 SSQLEHAKVTOTELMR-----ESFRQKQEBATESL-----KQOELE 1901

Qy 1375 TLAKIQESQSKQOSLNMKEDNETTKIVSEM-----EQFKPKOSALLRIEIMLGLSKRL 1430
 Db 1902 RLH--ESRARBEQALVELSKAEGVIDGYADEKTLFQIOEKTDIIDRLEQELLCSNRL 1959
 Qy 1431 QESHDEMSVAKKDDLQRLQELVLOSES-----DQLEKENIKEIVAKHLET----- 1475
 Db 1960 QELEAQOQIQEBRELLSRQKEAMKAEAGPVEOOLQOETEKIMKBKLEVOQCAEKVRRDDL 2019
 Qy 1476 EEBLVAHCCLBQEBETINELRVNLSEKETEISTIQKLEAINDKLQNK----- 1524
 Db 2020 QKQKALEIDVBSQVSRFIELE--QEKNTLMDLRQCNQALQKLEKMKFELDEQAIDR 2076
 Qy 1525 -----IQEIIYKEBQLNI-----KOISEVQENNELKQFKEHRKAKOSALQSIESKMLE 1573
 Db 2077 EHERDVQFQIQLEQKLVVPRFQIPSEHQ--TREVEQLANHLKEKTKDCKSELLSKQ 2134
 Qy 1574 LTNRLQESQEBE-----QIMIKEEMKRYQE-----ALOIE 1605
 Db 2135 LQDIOERNEEIEKLFPRVRELEQALLVSDATPQKVEDRKHFGAVEAKPELSLEVLQAE 2194
 Qy 1606 RQDKENTKEIVAKMKESQEKYQFLKMTAVNETQEKMCIEIHLKEQFTQK-----LNLE 1661
 Db 2195 RDAIDRKEKEIT-----NLEEQLEQFRE-----ELEKNNEEVQQLHMQLEIQKKESTRQ 2245
 Qy 1662 NIETENIRLTQILHENLEEMRSVTKERDRLRSVEETL-----KVERDQLKENL 1709
 Db 2246 ELEQEN-----KLFPQDDMEKLAGLAIKESDAMSTQDQHVLFQKFAIIQOEKEVEIDQLNEQV 2301
 Qy 1710 -----RETITRDLEKQEBELKIVHML--KEHQETIDKLRGIVSEKTN 1749
 Db 2302 TKLQOQLKITDNKVIIEKXNELIRDLETOELCMSQOEQCVKNREBIEQLNEVIEKLOQ 2361
 Qy 1750 EISNM-QK-----DLHNSDALKAQDLKIQBELRIAH----- 1780
 Db 2362 ELANIGQKTSMAVHLSSEADSLKHQDVVVIAEKLALEQOQVETANEEMTFMKNVLKETNP 2421
 Qy 1781 -----MHLKEQOETIDKLRGIVSEKTD--KLNSMQKDLNSNAKLEKIQLKANEH 1830
 Db 2422 KANQLTOELFSLKRRESVEKIQSIPENSVAIDHLSKDKPELVLTQ--DALKSLEN 2479
 Qy 1831 Q-----LITLKKOVNETQKVS-----EMEQLKKQIKQOS----- 1860
 Db 2480 QYFKSPENGKGSINLETRLLQLESTVSAKDLCTQYKIQKOMQEQOQFETEMLOKK 2539
 Qy 1861 -LTLKLEIENLNAQELHENLEEMKSVNK-----ERNLRRVEETLKLERDQ 1907
 Db 2540 IWNLOKIVEEKVAAALVSOIQLEAVQYAKFCODNQTTISSEPERTNIQNLQ--LREDE 2596
 Qy 1908 LKESQOETKARDLEIQO--ELKTARMLSKKHETVVDKLRKISKTIQISDIQKDLKS 1964
 Db 2597 LGSDISALTIRISELESQVEMHTSLILEKEQVEIAEK--NVLEKEKLELQKLEGN 2653
 Qy 1965 KDELQKTOELQKQELQRLVRKEDVNMHKK--INEMEQLKKQEPYVLCCEMDNFOL 2021
 Db 2654 EKKQREKKEKKSPODEVVLKTTTTELFSNBSGPFNELEALRAE--SVATKAELASY-- 2708
 Qy 2022 TKQLHESLEEIRIVAKE-----RDELRRIKESLK-MERDQOFIATLREMIARD 2067
 Db 2709 -KEXAKELQE--ELLVKETNMTSLQKDLQSOVDHLAAEKLSILEKEDETEVEQSKKAM 2766
 Qy 2068 RQNHQVPEKLLLS--DQOQHLMES-----LRBKCS--RIKELLKRYSE 2107
 Db 2767 FEPLPKLSKSIASOTDGTLLKISSNOTPQLLVKNAGIQINLQSECSEEVTEIISQFTE 2826
 Qy 2108 MDDHVECLNRLS-LDLE-KBIEFHRIMKLVLYSYVTKIEE-----QHECINKFEMD 2159
 Db 2827 KIEKQELHAAIILDMBSRHISSETITLKRHYVA--VOLLEKECGTLKAVIQCLRSKEGS 2884
 Qy 2160 FIDEVE-----KQKELL-----IKIQLHQQDCVPS--RELRLDLKLNMDLHI 2201
 Db 2885 SIPELAHSDAYQTEICSSDSGSDWGQGIYLTH--SQGFDIASEGRGESESATDSFPKKI 2943

QY 816 -----DOEFQNFKTLHMDPEQKYMVLEE-NERMNOBIV----- 848
 Db 1220 CEVNAEDKENS GDYISENDEPELQDYRYEQDFQENMHITLANKVTEYNLVLQTRLSK 1279
 QY 849 -----NLSKEAQF-----DSSIGAL-----KTELSYK 871
 Db 1280 INGOOTDGMKLEGEENLPEETEFLLSHSQMTNLEIDVNHKSKLSSLODLKTELEQ 1339
 QY 872 TOELOETREVBQRLNEMQLKEQLENRDSPLQTVREKTLITLKEQLTLEEVKTLTOEK 931
 Db 1340 VQELLESLSLQOQLKETE-----NYEAIHCLQRLQAVSESTVPPSLPDVDSVITE 1393
 QY 932 DDJLQOLQESQIERDQLKSDIHTVNNIDTQOLRNALRESLKHQHQFTINTLSKI----- 987
 Db 1394 SDAORTMYPGSCVKKINDGTIFESGFGVKEETNIVKLE--KQYQOLSEEVAKVISM 1451
 QY 988 -----SEEVSRNLHEENTGETKDEF-----QKWVGIDKKQOLEAKNTOTL-TADVK-- 1034
 Db 1452 STAFQOTELSRISGKENTASSQAHAVCQBOHVFNEKLSQDQIGFOTFTETVDVFKF 1511
 QY 1035 -----DNEIIEQQRKIFSLIOBKNELOQMLESVIAEKEOL-----KTOLKENIEM 1079
 Db 1512 BEFKPLSKELGHEKILLSNSPDHDPESKOCVLTISEMFSKDTFFVRSIHDEISV 1571
 QY 1080 -TIENQOEELRLIGDELK-KQOEIVAOEKHAIKKEGEL-----SRTCDRLAEVEEKLEK 1132
 Db 1572 SSNDASRQMLNBEQLEDMRQELVROQOE--QOATELLRQAHMRQWRQREDOEQLOEE 1629
 QY 1133 SOLOKQOQLLNQOE--MSEMQK-KINEIENLKN-----ELKN----- 1169
 Db 1630 IKRLNQLAQRASINDENLVSERVLLEBEALKQOLSLAGREKLCELENSSTQOTQNG 1689
 QY 1170 -----KELTLBHMETERL--ELAKL--NENVEEVKSTIKERKVLKELQKSF-----ETE 1215
 Db 1690 ENQGEVEEQTPREKELDRKPEDVPPILSNERYALQKANNRLKILLEVKTAAVEETI 1749
 QY 1216 RDHLRGYI-----RETEATGL--QTEELKIAHILKHEQETIDELRRSVEKTAQIINTQ 1269
 Db 1750 GRHVILGIDRRSKSOSASLIWRSEAAEASVKSVCHEHTRVTDE--SIPSYSGSDMPEN 1806
 QY 1270 DLE-----KSHTKLOEEI-----PVLHEEQELLPNVKVSQETQETWNE-LELLTQ 1314
 Db 1807 DINMSKVTEETELSORLVRSFGAGTEIDPNEELMLNIS--SRLQAAVEKLELAISET 1864
 QY 1315 STTKDSTTLARIEMLRLNEXFORBSQEBIKSLTKERDNLKTIKEALEVKHQLKEHIRE 1374
 Db 1865 SSQLEHAKVTOTELMR-----ESFRQKQATESL-----KCOEELRE 1901
 QY 1375 TLAKIQESQSQOEGLNKKENKNETTKIVSEM-----EQFKPKDSALLRIEIMLGLSKRL 1430
 Db 1902 RLH--EESRAREQLAVELSKAEGVIDGYADEKTLFERQIQTETDIIDRLQELLCASNRL 1959
 QY 1431 QESHDEMSVAKEKDDLORLQELVLOSES-----DOLKENIKEIVAKHLET----- 1475
 Db 1960 QELEAEQOQIEERELLGRKEMAKAEPVQQLQETELMKLEKJGVQCAEKVRDDL 2019
 QY 1476 BEELKVAHCLLKEQBETINELRVNLSEKETISTIQOLEALNDKLONK----- 1524
 Db 2020 QKQVKALBIDVEEQVSRFIELE--QEKNTLMDLRQOQOALEKQLEKQKRFLEQADIR 2076
 QY 1525 -----IQEIVEKEQLNI--KQISEVENNVNELKQFKHEKRAKDSALQSIESKMLE 1573
 Db 2077 EHERDVFOQEIQKLEQLKVVRFOPFISEHQ--TREVEQLANHLKEKTDKCSSELLSKQE 2134
 QY 1574 LTNRLQESQEBI-----QIMIKEKEMKRVQF-----ALQIE 1605
 Db 2135 LQRIQERNESIEKLEFPVRLEQALLVSADTFQKVEDRKHFAGAVEAKPELSLEVLQAE 2194
 QY 1606 RDQKENTKIVAKMBSQEKQYOFKMTAVNETQEKMCETIEHLKEQETOK-----LNLE 1661
 Db 2195 RDAIDRKEKEIT--NLEEQLEQFRE-----ELENKNEEVQQLHMQLEIQKKESTRLQ 2245
 QY 1662 NIETENRLTOILHENLEEMRSVTKERDDLSVEETL-----KVERDQLKENL 1709

Db 2246 ELBQEN-----KLFKDDMEKLGIAKESDAMSTQDQHVLFKGFQAQIIQKEVEIDQLNEQV 2301
 QY 1710 -----RETITRDLKQOEELKIVMHML--KEHQETIDKLRGIVSEKTN 1749
 Db 2302 TKLQOQLKITTDNKVIEEKNEILRDLETOIECLMSDOECVKRRREEEIEQLNEVIEKLOQ 2361
 QY 1750 EISNM-QK-----DLEHNDALKQAQDLKIQEELRIAH----- 1780
 Db 2362 ELANIQKTSMNAHSISEEADSLKHQLDVVVIAEKLALEQQVETANEEMTFMKNVLKETNF 2421
 QY 1781 -----MHLKEQOETIDKLRGIVSEKTD-KLSNMQKDLNSNAKQEKIOELKANEH 1830
 Db 2422 KMQLTQELFSLAKRESSEVEKIOSIPENSNNVAIDHLSDKQKPEVLVITE--DALKSLEN 2479
 QY 1831 Q-----LITLKKDVNETOKTKVS-----EMEQLKKQIKQDS----- 1860
 Db 2480 QTVFKSFEENGKSGIINLETRLLOLESTVSAKDLELTQCYKQIKDMQEQOQOPETEMLOKK 2539
 QY 1861 -LTLKLEIENLNLAEHLHENLEEMKSVMK-----ERDNLRRVEETLKLERDQ 1907
 Db 2540 IVNLQKIVEEKVAAALVSOILEAVQYAKFCODNOTISSEPERTNIQNLQ--LREDE 2596
 QY 1908 LKESLOETKARDLEIQO-----ELKTAEMLSKEHKEVTDKLRKESKTIQISDIOKOLDKS 1964
 Db 2597 LQSDISALTIRISELESQVVEHMTSLILKEQVETIAEK--NVLKEKELKLELQKLEGN 2653
 QY 1965 KDELQKKIOELQKHELQLLRVKEDVNMHKK--INEMEQLKQKQEPNVLCKCEMDFQL 2021
 Db 2654 EKKQREKERRSPQDVEVLKTTTELPHSNEESGFFNELRALRAE--SVATKAEIASY-- 2708
 QY 2022 TKKLHLESLEIRIVAKE-----RULRRIKESLK-MERDQFIATLREMIARD 2067
 Db 2709 -KEKAELQOE-ELLVKETNMTSLQKDSQVRDLAEAKEKLSILEKEDETEVEQESKACM 2766
 QY 2068 RQNHQVKEPKRLLS--DGOQHLMES-----LREKCS--RIKELLKRYSE 2107
 Db 2767 FEPFLKPKSKSTASQTDGTLKISSNQTPQIILVKNAGIQINLOSCSEVEETIISQFTE 2826
 QY 2108 MDHYECINRLS-LDLE-KEIFHFRIMKKLYLSVTVTKIEE-----QHECINKFEMD 2159
 Db 2827 KIEKQELHAAEILDMSRHSISETLKEHVVA--VQLLKEECGTLKAVIOCLSKESG 2884
 QY 2160 FIDEVE-----KQKELL-----IKIOHQDQDCDVPS--RELRLKLNQNMDLHI 2201
 Db 2885 SIPELAHSDAYOTREICSSDSGSDWGQGIYLTH--SQGFDIASEGRGESESATDSFPKKI 2943
 QY 2202 BEILKDFESEFPSTKTEFQOVL-----NPKEMTOFLEEWLNTF--FDIEKLK 2248
 Db 2944 KGLLR-----AVHNEGQVLSLUTESPYSQDGHSTIQVSEPPWLEERKAYINTISSLK 2995
 QY 2249 NGIOK-ENDRICQVNN-----FFNNRIIAIMNESTE----- 2278
 Db 2996 DLITKQOLQREAEVVDSSQSHESFSDWREGLLALQVFLERSVLAAFTTELTAQT 3055
 QY 2279 -----FERSATISKEVEDLKLKERNK-LFKNYOTLKTSLASCAQVNPPTQDNK 2329
 Db 3056 DAVGLLNCLEQIOEQGVYQYQAMECLQKADRSILSIEIQALHAQMGCRKITLKEQESE 3115
 QY 2330 NP-----HVTSRAT-----OLITTEK--IRELENSLHEAK-- 2356
 Db 3116 KPSOELLEYNIOQKQOSQMLEMQVELSNKDRATELOELQESKEMVVAELKSELATQKLEL 3175
 QY 2357 ---ESAMHKEKSKIIKMOK-----ELEVNDIIAKLQAKVHESNKCLEKTEIQLVQL 2405
 Db 3176 ETTLKAQHKLKELEAFLEVKDKTDEVHLLNDTLASEQKSRQELQWALEKEKA----- 3239
 QY 2406 DKVALGAKPYKEETIDLMKLVKIDLEKMKNAKEFEKISATKATVEYQKEVIRLLREN 2465
 Db 3230 -KLRSEERDEKEELEDLKFSL--ESQKQNLQ-----LNLLEEQKQLNESQOKI 3277
 QY 2466 RRSQQAODTSVISEHTDPQPSNPKPLTCGGSGSIVQNTKALILKSEHRLRKEISKLKQON 2525

Db	3278	ESQRMIDYDLGSEE---QGRNLEL-----QVLLSEKVRIREMSSTLDRER	3320	Db	712	SKLKDQQSLVNSKSEMTLQINELQKEIE-----ILRQE-EKEKGTLEQVEQEL---	760
Qy	2526	R-----QLIKOKNELLSNNQHLNSNEVKTWKERTLK-----	2555	Qy	514	YFQLRTEKEEMELKAKENDLDE-----	536
Db	3321	ELHAQLQSSDGTQOSRPPLPSEDLLKELQOLEEKHSRIVELLNETEKYKLDLSQTRQOM	3380	Db	761	--QLKTELLEKQWKEKE--NDLQEKFAQLEAENSILKDEKKTLEDMLKIHTPVSQEERLIF	817
Qy	2556	---REAHKQVTCENSPKSPKVTGTASKKKQITPSQCKERNLQDPVPKSPKSCFFPDSRSK	2612	Qy	537	FEALERKTK-----KQOEMOLIHEISNLKLV-----KHREVVYNQDLE	574
Db	3381	EKDRQVHRKTL-----QTEQANTEGOKK-MHELOSKVEDLQOLEEKQOVVKLDLEGQ	3434	Db	818	LDSIKSKSDSVMEKIBILIEENEDLKQOCIOLENEIEKQRTTFSAEKNFVNYQELQ	877
Qy	2613	SLSPHPVRFYFDSNLSGLCEPQVAGAESVD-----SQCPFWAASGK	2655	Qy	575	NELSSSVELLREKEDQIKKLOEYIDSOKELENIKMDLSYSLESIEDPK-QMKQTLFD---	629
Db	3435	RLQ-----GIMQEFQKQELEREKESRRILYQNLNPTTSLTSDR	3476	Db	878	EETACLLKVKDDLEDS-KNQOELEYKSKALANEEL--HLQRIINPTTVKMKSSVDFDRDKT	934
RESULT 12							
US-10-370-685-100							
; Sequence 100, Application US/10370685							
; Publication No. US20030215903A1							
; GENERAL INFORMATION:							
; APPLICANT: Hyman, Paul							
; TITLE OF INVENTION: Nanostructures Containing PNA Joining and Functional Elements							
; FILE REFERENCE: NANF.P-004							
; CURRENT APPLICATION NUMBER: US/10/370,685							
; CURRENT FILING DATE: 2003-02-21							
; PRIOR APPLICATION NUMBER: 10/080,608							
; PRIOR FILING DATE: 2002-02-21							
; NUMBER OF SEQ ID NOS: 159							
; SOFTWARE: PatentIn version 3.2							
; SEQ ID NO 100							
; LENGTH: 3911							
; TYPE: PRT							
; ORGANISM: human							
US-10-370-685-100							
Query Match							
Best Local Similarity 19.5%; Score 1055; DB 15; Length 3911;							
Matches 665; Conservative 616; Mismatches 1113; Indels 1016; Gaps 145;							
Qy	26	TAQVYKTDNNVIYQVDSKSFNDFRVPHGN-----ETTKVYERHTAAPILDSATQGYNGT	81	Db	1292	IWGQOTDGMKLEFGENLPKEETBFLSIHSQMTNLEDIDVNHKSKLSLQDLEKTKLSEQ	1351
Db	307	TWOISFLQEKIKVYEMEQDK-----KVENSKEIEQEKETIEELNTXIIIE-----	353	Qy	872	TOELOEKTRFVOERLNEQOLKEQLENRDSPLQTVREKTLITLQOTLEEVKTLQOR	931
Qy	82	IFAYQOTASGKTYTMWG-----SEDLGVIPRAIHDFQIKKFPDPREFLLRVSVMEIYN-	136	Db	1352	VQELESLSLQQLKETEQ-----NYEAIHCLQKELQAVSESTVPSLPVDSVITE	1405
Db	354	-----EKTLLEKDKLTADKLGLGELQEOIVQKNQEIKNMK-----LELINS	395	Qy	932	DDLQOLQESLQIERDQKSDIHDVTYNNMIDTQEOQLRNALESIKHQHQTINTLUSKI---	987
Qy	137	-----ETITDLICGTQKMKPLIREDVNRNVYVADLTEEVVYVTSSEWALKWITKGEKS	188	Db	1406	SDAQRTYPGSCVKKNIDGTIEFSGEFGVKEETNIVKLE--KQYQEOLEEEVAVIVSM	1463
Db	396	KOKERQSSEIEKQLMGTVVEELQ-----KRNHKDSQFETDIVQMEQETORKLEQL-RAELD	450	Qy	988	-----SBEVSRNLHMBENTGETKDEF-----QOKMVGIDKKQDLEAKNTOTL-TADVK--	1034
Qy	189	RHYGE--TKMQRSSRSHITFMILLESREKGEPSNC-----EGSVKYSHLNLVDL	236	Db	1464	SIATAQQTSLRSISGCKENTASSKQAHAVCOQEOHYFNEMKLSQDQIGQETFTVDVVKF	1523
Db	451	EMYGQOIVQMKQLIRQMAQMEEMKTRHKGEMENALRSYSNITNEDQIKLMNVAINEL	510	Qy	1035	-----DNEIITBQKFIKSLIOEKNELOQMLESVIAEKEQL-----KTDLKENTBM	1079
Qy	237	-----AGSERAQATGAAGVRLEKGCNINRSLFILGOVLIKSDG-----QVGGFTNYRD	285	Db	1524	EEFKPLSKELGHEGHEILLNSDPDHDIPESKDCVLTISEMFSKDKTFIVRQSIHDEISV	1583
Db	511	NTKLODTSNQKELKEELGLILBEKCALQOQ--LEDIVELSFSEEQIQRARQITAEQE	567	Qy	1080	-TIENQOEELRLGLDELK-KQBEIVAQEKNHAIKKGEL-----SRTCDRLAEVEEKLEKX	1132
Qy	286	SKLTRILQNSLGNPKTRICTITP-----VSFDETLTALQFASTAKYMNTPYNEVST	340	Db	1584	SSMDASRQMLMNEEQLEDMRQELVQYQEH--QOATELLRQAHMRQMERQEDQEOLEB	1641
Db	568	SKL-----NEAHSLSLTVEDLKBAIVSASESRKELE-----LKRAEAVTNYKI	610	Qy	1133	SOOLQEKQOOLLNVQEE--MSEMQK-KINEIENLKN-----ELKN-----	1169
Qy	341	DEALLKRYKEIMDLKQLEEVSL-----TFAQAMEKDQLAQLLEKDLQKQVNEKIENL	397	Db	1642	IKRLNRLQARSIDNENLVSRERVLLEELKALKQSLAGREKLCCELNRNSSTQONGN	1701
Db	611	KLEMLEKEKNALDRMAESQEAELERLTOLLFSHEEELSCLKEDLEIEHRINIEK-----	666	Qy	1170	-----KELTSHMEPERL--ELAQKL--NEWYEVKSTTKERKVLKEQKSF-----ETE	1215
Qy	398	TRMLVTSSSLITLQOELKAKRKRRVTWCLGKINKKNSYNAQDQFNIPNTITTKTHK--LSI	455	Db	1702	ENQGEVEEQTFFKEKLDKRPEDVPPILSNERYALQKANNRLKILLEVVKTTAAVEETI	1761
Db	667	-----LKNLGIHYKQIID--GLQNEWSQKIETMQPE-KDNLITKQNQLILEI	711	Qy	1216	RDLRGYI-----REIETGL--QTKBELKIAHILKHEQETIDELERSVSEKTAQINTQ	1269
Qy	456	NLLREIDES--VCSESDFVFSNTLDTL--SEIETWNPATKLLNQENITESELNSLRADYDNLVDL	513				

Db 1762 GRHVGLDRSSKQSGASLIWRSEAEASVKSVCVHEEHRTVTDE---SIPSYSGSDMPRN 1818
Qy 1270 DLE-----KSHTKLOBEI-----PVLHEOEQLPNVKVSETORTWNE--LELLTEQ 1314
Db 1819 DINMSKVTBEGTSLQRLVRSFGAGTETIPENELMNIS--SRQAQAEKLEJAEIS 1876
Qy 1315 STYKDSITLARIEMERLRLNEKPOESQREIKSLTKERDNLTKIKAELVHDQJKEHIRE 1374
Db 1877 SSQLEHAKVTQTELMR-----ESFRQKQEAESL-----KQOEELRE 1913
Qy 1375 TLAKIQESQSKOESQKLMNEKNETTKIVSEM-----EOPFKDSALLRIHEMLGLSKRL 1430
Db 1914 RLH--EESRAREQAVLSKAEGVIDGYADEKTLFERQIQEKTIDIDRLEQELLCASNRL 1971
Qy 1431 QESHDEMSVAKEDDQLRLQELVLSSES-----DOLKENIKIEIVAKHLET----- 1475
Db 1972 QELEAEQOQIQEERELLRSRQKAEAGFVEQQLLOETEKLMKEKLEVOQCAEKVRDDL 2031
Qy 1476 EBEELKVAHCCLEQOETINELRVNLSEKTEISTIQOLRAINDKLQNK----- 1524
Db 2032 QOVKKALEIDVEQVSRFIELE---QEKNTLMDLRQONQALEKQLEKMKRKFLEQALDR 2088
Qy 1525 -----IORIYKEBQLNT-----KOISEVOENVNELKQFKEHRKAKDSALQIESKMLE 1573
Db 2089 EHERDVFOQEIQKLEQOLKVVPFQIPSEHQ---TREVEQLANHLKGTDKCSELLLSKEQ 2146
Qy 1574 LTNRLQESQBEI-----QIMIKEKEEMKRVQE-----ALQIE 1605
Db 2147 LORDIQERNEETKLEFRVRELQALLVSADTPQKVEDRGHFGAVEAKPELSLEVQLQAE 2206
Qy 1606 RDOLEKENTIEIVAKMESQEKQVQPKMTAVNETQBMKRIEHLKBOFTQK-----LNLE 1661
Db 2207 ROAIRDKKEIT-----NLEEQLEQPRE-----ELENKNEBEVQOLHMOLEIQKKESTTRLQ 2257
Qy 1662 NIETENIRLTQILHENLEBMSVTKERDDLRSEVETL-----KVERDQKLENL 1709
Db 2258 ELEQEN---KLFKODMEKLGAKESDAMSTQDQHVLFKGAQIITOEKEVEIDQLNEQV 2313
Qy 1710 -----RETITRDLKQELKIVHMHIL--KEHOETIDKLKGIYSEKTN 1749
Db 2314 TKLOOQLKITTONKVIIEKNELIRDLTQIECLMSQECVKRNEHEIEQLAIEVEKLOQ 2373
Qy 1750 ETSNM-QK-----DLEHSDALKAQDLKIQEELRIAH----- 1780
Db 2374 ELANIGQKTSMAHLSSEADSLKHQLDVVIKALEQOVETANEEMTFMKNVLKETWF 2433
Qy 1781 -----MHLKEQOETIDKLKGIYSEKTD--KLSNMQKDLNSNAKLOEKLOELKANEH 1830
Db 2434 KMNQLTQELFSLKREESVEKIQSIPENSUNVAIDHLSKDKPELEVVLITE--DALKSLEN 2491
Qy 1831 Q-----LITLKQOVNETQKVS-----EMEQKKQIKDQS----- 1860
Db 2492 QTYFKSFBENGKSGIINLETRLLQLESTVSAKDLELTQCYKQIKDMQEQGPETEMLOKK 2551
Qy 1861 -LTLKSLIENLNAQELHENLEEMKSVNK-----ERNLARVEETLKLERDQ 1907
Db 2552 IVNLOKIVEEKVAAALVSOIQLEAVOYAKFCQDNQTISSEPERTNIQNLQ--LREDE 2608
Qy 1908 LKESIQETKARDLEIQ--ELKTARMLSKKHETVDKLRKIKSEKTIQISDIQKLDKS 1964
Db 2609 LGSDISALTALISELSQVEMHTSLILEKEQVEIAEK--NVLEKEKKLLELQKLEGN 2665
Qy 1965 KDELOKKTQELQKLEQLLRVKEDVNMGHKK---INEMEQLKQPEPNYLCKCEMDNFOL 2021
Db 2666 EKKQREKKGKSPQDVEVLKTTTELPHSNEESGFNLEALRAE---SVATKAELASY-- 2720
Qy 2022 TKLHLESLEERIVAKE-----RDELARIKESLK-MERDOFIATLREMIARD 2067
Db 2721 -KEKAELQOE-ELLVKETNMWTSLOKDLQSVRDHLAAEAKELSLILEKEDETEVEQSKKACM 2778
Qy 2068 RQNHQVPEKELLSDGQOHLMESLREKCSRIKELLKRYSEMDDHVECLNRLSLDLKEKIE 2127
Db 2779 FEPLPIKLSKIASQTDGTLKISSNQTPQI--LVKNAG-----IQINLOSECS 2825

Qy 2128 FHRIMKKLVLSVTVTKIKEBQEHCEINKFEMDFIDEVEK-QKELLIKIQLHQDQCVPSR 2186
Db 2826 SEEVETIISQFTEKIERMQELHAAEILDMESRHSISETETLKRHHVAVQLLKEBQGLTKA 2885
Qy 2187 ELRDKL-----NONMDLHIEEILKDFSES---EFP 2214
Db 2886 VIQCLRSKVFVFNMCFTLCSGSDWGQGIIVTHSQGFDI-ASEGRGESESATDSFP 2944
Qy 2215 -----SIKTEPQQVLS-----NRKEMTOFLEEWLNT--FDEKLKNGIQK 2253
Db 2945 KTIKGLLRVAHNEGMQVLSLTPESYSGEDHSIQOVSFPEWLEKAKYINISSLKLITK 3004
Qy 2254 -ENDRICQVNN-----FFNRIIIMNNESTE----- 2278
Db 3005 MQLQREAEVYDSSQSHESFSDWRGELLALQOVFLERSVLLAAFRTELTAFTGTTDAVGL 3064
Qy 2279 ---FEERSATISKWEODLKSLEKNEK-LFKNYQTLKTSLSAGAQVNPPTQKNP--- 2331
Db 3065 LNCLEQRIQEOGVYQAMECLQKADRRSLLSIOLHAQMGKRTILKREQSEKPSQE 3124
Qy 2332 -----HVTSRAT-----OLITTEK--IRELENSLHEAK-----ES 2358
Db 3125 LLEYNIQKQSQMQLMQLVSSMKDRATELOEQLSSEKQVVAELKSELATQKLEETTLK 3184
Qy 2359 AMHESKILIMQK-----ELEVTNDIIAKLOAKVHESNCKLEKTKETIQVLODKVAL 2410
Db 3185 AQHKLHELEAFRLVCKDTDEVHLNDTLASEQKKSRELQWALEKEKA-----KLGR 3237
Qy 2411 GAKPYKEETEDLKMVLKIDLEKMKNAKEFEKEISATKATVEQKEVIRLLRENLRSSQ 2470
Db 3238 SEERDKEELEDLKFSL---ESQKQRLQ-----LNLLEQOQKOLLNESQOKTESQRM 3286
Qy 2471 AOTSVISHTDPPSNKPLTCGGSGIVQNTKALIKSEHIRLEKEIKSLKQONE--- 2526
Db 3287 LYDAQLSSE---QGRNLEL-----QVLESEKVRIREMSSTLDRELRHAQ 3329
Qy 2527 -----QLIKQKELLNNQHLNNEVKTWERTLK-----RE 2557
Db 3330 LOSSDGTQSRPLPSBLLKELQLEKHSRIVELLNETEKYKLDLSLQTRQOMEKDRQ 3389
Qy 2558 AHQVTCENSPKSPKVTGTASKKKOITPSQCKERNLODPVKESPKSCFFDSRSKSLPSP 2617
Db 3390 VHRKTL-----QTEQANTEGQKK--MHELOSKVEDLOQLEEKQOVYKLDLEGORLQ-- 3441
Qy 2618 HPVRYFDNSSLGLCPVEQVQAGASYD-----SQGPWHASSGK 2655
Db 3442 -----GIMQEFQKQELEREKESRRILYQNLNEPTTWLSLTSR 3480

RESULT 13

US-10-408-765A-1839

; Sequence 1839, Application US/10408765A

; Publication No. US20040101874A1

; GENERAL INFORMATION:

; APPLICANT: Ghosh, Soumitra S.

; APPLICANT: Fahy, Eoin D.

; APPLICANT: Zhang, Bing

; APPLICANT: Gibson, Bradford W.

; APPLICANT: Taylor, Steven W.

; APPLICANT: Glenn, Gary M.

; APPLICANT: Watcock, Dale E.

; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION

; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME

; FILE REFERENCE: 660088.465

; CURRENT APPLICATION NUMBER: US/10/408,765A

; CURRENT FILING DATE: 2003-04-04

; NUMBER OF SEQ ID NOS: 3077

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 1839

; LENGTH: 3911

; TYPE: PRT

; ORGANISM: Homo sapiens

Fri Aug 6 10:49:04 2004

US-10-408-765A-1839

Query Match 7.9%; Score 1055; DB 16; Length 3911;
Best Local Similarity 19.5%; Pred. No. 5.4e-44; Indels 1016; Gaps 145;
Matches 665; Conservative 616; Mismatches 1113;

QY 26 TAQVYKTDNNVYQVDSKSFNDRVFHGN-----ETTNNVYBEIAAPIIDSALIOGYNGT 81
DB 307 TMOISFLQEKIKYEMEQDK-----KVNSNKEIEQEKETIIEBELNKLIEE----- 353
QY 82 IFAYGQTASGKTYTMMG-----SEDLGLVPIRAIHDFQKIKKPPDRFLLRVYMEIYN- 136
DB 354 -----EKKTLKDKLTADKLLGELQEQIVQKNQIKNMK-----LELTNS 395
QY 137 -----ETITDLCTQKMKPLIREDVNRNVYVADLTVEEVYITSEMALKWITKGEKS 188
DB 396 KOKEROSSEBIKOLMGTVBELQ-----KRHKDQOFETDIVORMEQETORKLEQI-RAELD 450
QY 189 RHYGE--TKNORSSRSHTIPRMILESGREKEGPSNC-----EGSVKVSHLNLVDL 236
DB 451 EMTGQOIVOMQOELIROHMAQMEEMKTRHKEMENALRSYSNITVNEDOIKLMMVAINEL 510
QY 237 -----AGSERAQTAGAAGVRLKGCNINRSLFQGVVKKLSDG-----QVGGFINYRD 285
DB 511 NIKLQDTSQKELKEELGLILEKCALQO---LEDLVEELSFRBQIQARQTIABQE 567
QY 286 SKLTRILONSLGGNPKTRIICTIP-----VSFDETLTALQFASTAKYMKNTPVVNEVST 340
DB 568 SKL-----NEAHKSLSTVEDLKABIVSASESRKELE-----LKHEAEVTNYKI 610
QY 341 DEALLKRYKEMIDKKQLEBVSLE--TRAQAMEKQDQAQLLEKXOLLQKVQNEKIENL 397
DB 611 KLEMLEKKNVLDORMAESQABELRLTQLFSHEBELSKLEDLEIHRINIEK----- 666
QY 398 TRMLVTSSSLTQOELKAKRRVTCVCGINKNMKNVYAOFPNIPNITTKHK--LSI 455
DB 667 -----LKNLGHYKQOID---GLONEMSQKLETWQFE-KDNLLITKQOQLILEI 711
QY 456 NLLREIDES-VCSESDVSFNTLDTI-SEIWNPAKLANQENIESELNLRADYDNLVLD 513
DB 712 SKLDLOQSLNSKSEEMTLQINELQKEIE-----ILRQE--EKEKGTLSQEVQEL--- 760
QY 514 YEQLRTEKEEMELKKEKNDLDE-----NDLOKEFPAQLAENSILKDEKTKTLEDMLKIHTPVSOBERLIF 817
DB 537 FEALERKTK-----KQOEMQLIHEISNLKXV-----KHREVVYNODLE 574
QY 818 LDSIKSKSDSVWEKEIETIIBENEDLKQCCIQLNEEIEKORNTFSFAEKNFEVNYQELQ 877
QY 575 NELSSKVELLREKEDQIKKQLEVIDSQKLENTKMDLSYSLIESIEDPK-QMKQTLFD--- 629
DB 878 EYACLLKVKDLEDS-KNKQELVYSKULKALNEEL--HLQRIPTVVKMSVDFDEK 934
QY 630 --AETVAL-----DARKESAFILSENLEKMKELATYIKOMENDIOLYQSLEA 678
DB 935 FVAETILEMEVVEKOTTELMEKLEVTREKLELSQRLSDLSQOLKQKHGEISFLNEEVKS 994
QY 679 KK--KXQVDLE-KELOSARNE-----ITKLTSLIDGKVPKDLNLELEGKTTDLQ 726
DB 995 LKQKEQVSLRCRELEIINENRAENVQCDTQVSLLDGW---TWTSGRAGSVSKYN 1051
QY 727 KEINKE---VRE-----NEALREVIILSELKSL-----PSEVERLRKEI 763
DB 1052 KSGFGEKIMVEDKVSFENMTVGEESKQOQLIIDLHLPSTVKESSLRATQPSENDKLOKEL 1111
QY 764 QD-KSEE-----LHI-----ITSEKDK-----LFSEVH-KESRVQ 792
DB 1112 NVLKSQNDLRLQMEAGRICLSIVYTHVDQVREYMEKDKALCSLKEELIPAEQEKIK 1171
QY 793 GL-----LE-----EIKTKDDLATQTSNYKST----- 815
DB 1172 ELQKIHOLELQTKMTQETGEGKPLHLILIGLQKAVSECSYFLQTLCSVLGEVYTPALK 1231

QY 816 -----DQBFQNFKTLHMDPEQYKMLVBE-NERMNOEIV----- 848
DB 1232 CEVNAEDKENSQDYISENEDPELQDYRYEVQDQENHLLNKVTEYKLLVLQTLRLSK 1291
QY 849 -----NLSKEAQKF-----DSSIGAL-----KTELISYK 871
DB 1292 IMGQQTDMKLEPGEENLPKEETEFLSIHSQMTNLEDIDVNHKSLSLSLQLEKTKLEEQ 1351
QY 872 TQLOEKTRREVQRLNEMEQLEQLENRDSPLQTVREKTLITITEKLOOTLEEVKTLTQEK 931
DB 1352 VQELSELSLSLQOQLKETEQ-----NYEABIHCLQKRLQAVSESTVPPSLPVDVSVITE 1405
QY 932 DDLKQLOESQIERNQKSDIHDTVMNIDTQOLRNALLESKQHOETINTLSKI----- 987
DB 1406 SDAQRTMYPGSCVKKNIDGTIEFSGFVKEETNVVKLE--KQYQOLEEVAKVIVSM 1463
QY 988 -----SEVSRNLHMEENTGETKDEF-----QOKMVGIDKKODLEAKNTOTL-TADVK-- 1034
DB 1464 SIAFAQOTELSRISGGKENTASSQOAHAVCQEQHYFNEMKLSQDQIGFOTFETVDVKFK 1523
QY 1035 -----DNEITIROOKIFSLIOEKNELOQMLESVIAKEQOL-----KTDLENLEM 1079
DB 1524 EEPKPLSKELGEGHKEILLSNDPHDIPESKDCVLITISEEMFSKDKTFIVRQSIDHLELSV 1583
QY 1080 -TIENQOEELRLGDELK-KQOEIVAOEKNHAIKKEGEL-----SRTCDRLAEVEEKLEK 1132
DB 1584 SSMDASQMLNEEQLEDMQELVRQYQEH--QOATELLRQAHNRQMERQEDQEQLE 1641
QY 1133 SQLOEQKQOQLNVQEE--MSEMOK-KINETENLKN-----ELKN----- 1169
DB 1642 IKRLNRQLAQRSSIDNENLVSERERVILLELEALKQLSLAGREKLCCELRNSSTQONGN 1701
QY 1170 -----KELTLEHMETERL--ELAQL--NENYEEVKSITKERVKLEKQSF-----ETE 1215
DB 1702 ENQGEVEEOTPKXELDRKPEDVPPFELLSNRYALQKANNLLKILLEVNVKTTAAVESTI 1761
QY 1216 RDHLRGYI-----REIATGL--QKKEUKIAHIHLKHOETIDELRRVSEKTAQIINTQ 1269
DB 1762 GRHVILGILDRSKSQSSASLIWRSEAEASVSKVHEHTRVTDE--SIPSYSGMDMPN 1818
QY 1270 DLE-----KSHTKQEBEI-----PVLHEOELLPNVKKVSTQETMAE-LELLTEQ 1314
DB 1819 DINMWSKVTEBGTLSQRLVRSFAGTAGTEIDPENELMLNIS--SRLOAAVEKLELAIS 1876
QY 1315 STTKDSTTLARIEMERLRLNEKFBQESQEBISLTKERDNLKTIKEALEVKHDLQKEHIRE 1374
DB 1877 SSQLEHAKVTQTELMR-----ESFRQKQEBATESL-----KCOEELRE 1913
QY 1375 TLAKTQESQKQOESQSLNMKKNEDNTEKIVSEM---EQFKPKDSALLARIEMGLSKEL 1430
DB 1914 RLH--EESRAREQLAVELSKAEGVIDGYADEKTLFERQIQEKTDIIDRLQELLCASNL 1971
QY 1431 QESHDEMSVAKEKDDQLQRLQEVLOSES-----DQKENIKEIVAKHLET----- 1475
DB 1972 QELEAEQQIQIQRERELLSSKQEMAKAEPVQQLQETEKLMKEKLEVQCOAKVRRDL 2031
QY 1476 ERELKVAHCLLKEQETINELAVNLSEKETETISTIQKOLEAINDKLONK----- 1524
DB 2032 QKQVKALEIDVBEQVSRFIELE--QEKNTLEMDLDRQQQALEKQLEKWKPFIDEQAIDR 2088
QY 1525 -----IQEIVKEQEQLNI-----KOISEVQENVVELKQFKERKAKOSALOSIESKMLE 1573
DB 2089 EHRDVFQOEIQKLEQOLKVVPRFQIPSEHQ--TREVEQLANHLKEKTKDCSELLSKQE 2146
QY 1574 LTNRLQESQEBEI-----QIMKEKEMKRVQE-----ALQIE 1605
DB 2147 LQORDIQERNEEIEKLEFRVRELEQALLVSADTFQKVEDRKHFGAVAKPELSLEVLQAE 2206
QY 1606 RDOLKENTKEIVAKMESQEKQYFLKMTAVNETQEKMKCEIEHLKEQFETQK-----LNLE 1661
DB 2207 ROAIDRKEKIT-----NLREOLEQFRE-----ELENKNEEVQQLHMQLEOKKESTRLQ 2257

Db	712	SKLQOLQSVNSKSEMTLOINELQKEIE-----ILRQE-EKEKGTLEQVQEL---	760
Qy	514	YEQLRTEKEEMELKLEKNDLDE-----	536
Db	761	--OLKTLELLEKQMKKE--NDLOEKFAQLEAENSILKDEKKTLEMDLKIHTPVSQERELIF	817
Qy	537	FEALERKTK-----KQEMQLIHEISNLKLV-----KHREVVYNQDLE	574
Db	818	LUSIKSKSDSVWEKIEIILIBENEDLKQOCIQNEEIEKQRTTFSPAENFVNVQELQ	877
Qy	575	NELSSKVELLREKEDQIKKLOEYIDSKLENIKMDLSYSLESTEDPK-QMKQLFD----	629
Db	878	EYVACLKVKVDLEDS-KNQELEYKSLKALNEEL--HLQRINPTTVKMSVFDDEKT	934
Qy	630	--AETVAL-----DAKREAFLESENLEKEMKELATTYKQOMENDIQLYQSOLEA	678
Db	935	FVAETLEMGVFEKDTTELMEKLEVTREKLELSORLSDLSBQKQKHGEISFLNEEVS	994
Qy	679	KK--KMOVDLE-KELQAFNE-----ITKITSIDGKVPKDLLCNLEBKITDLO	726
Db	995	LKEKEQVSRCRELEIIINHNAENVQSCDTQVSSLLDGVV--TWTSRGAEGSVSKVN	1051
Qy	727	KELNKE-----VEE-----NEALREBVILLSKSL-----PSEVERIRKEI	763
Db	1052	KSFGEESKIMVEDKVSFENMTVGBESKQEQIILDLPLSVTKESSLRATOPSENDKLOK--	1109
Qy	764	QDKSEELHITSEKDKLFSEVVHVESVQGLLEIGTKDLDATTQSNYKSTQDEFQNFK	823
Db	1110	-----ELNVLSKQNDL-----RLQMAQRI-----CLSLVYST-----	1138
Qy	824	TLHMFQOKYKWLNEERNMQEIVNLSKEAQKFDSSIGALKTSLYKTOBLOKTRVQ	883
Db	1139	--HVDQVREYM-----ENK-----DKALCSLKEELIPAQ-----EEKIKELQ	1174
Qy	884	ERANEMQLKEQLENRDSPLQTVREKT-----LITEKLOQTLLE-----VKTLTOE	930
Db	1175	KIHOLE-----LQTMKTQETGDEGKPLHLIGLOKAVSECSYFIQTLCSV	1221
Qy	931	KDDLKQLOESLOIERDQKSDIHDTVMNIDTQEQRLNALSLLKQHOETINTLSKISSE	990
Db	1222	LGYYTTPALKCEVNAED--KENSQDYISENED--PELQDYREYVQDFQENMHTLLNKVTEE	1278
Qy	991	VSRNLHM-----PENTGETKDEFQQRMGVIGDKODLEAKNTOTLT	1030
Db	1279	YKLLVLQTRLSKINGQOTDGMKLEFGEENLPKBEFELSHTSOMTNLEDIVNHSKLS	1338
Qy	1031	A--DVKDNEITQOKKIPSLI-----QEKNEILOOMLES-----	1061
Db	1339	SLQDLEKTKLEFQVQELESLISSLOOQKETEQNYEAIEHCLQKQLQAVSESTVPPSLPV	1398
Qy	1062	--VIAKEBOLKT-----DLKENIEMTIE-----NOBEL-----	1087
Db	1399	DSVWITESDAORTMYPGCVKKNIDGTITFSGEFGVBKBTNIVKLEKQYQOLSEEVAK	1458
Qy	1088	-----RLIGDE-----LKKQOBIVAQEKNAHAIKGBEISRTCDRIA----	1123
Db	1459	VIVMSIAFAQOTELSRISGGKENTASSKQAHAVQOEQHYF-----NEMKLSQDQIGFQTF	1515
Qy	1124	EVEFKLEKKSQOLO-----EKQOOLINVOEEMSEMOKKINITE	1161
Db	1516	ETVDVKFEEFKPSLKGELGEGHKEILLNSDDPHDIPESKDCVLTISEMPSKDKTTFIVRQ	1575
Qy	1162	NLKNELK-----NKELTLEHMETELELAQKLNENYEVKSITTKERKVLKELQKSFETE	1215
Db	1576	SHIDEISVSMOASRQMLNNEOLE--DMRQELVRQYQEHQATQRSSIDNE--NLVSE	1630
Qy	1216	ROHLRGYIREIATGLQTKKEIKIAHIEKHQETIDELRRSV-----SEKTAQIINTQ	1269
Db	1631	RE--RVILLELEA-----LKQLSLAGREKLCCELRNSTQTONGNENQGEVEEQT	1678
Qy	1270	DLKSHTKLOEETP--VLHEEQE-----LLPNVKVSETQETMNE--LELLTE	1313
Db	1679	FXEKELDRKPEDVPPPEILSNERYALQKANNRLIKILLEVVVKTAAVEETIGRHVLGILDR	1738
Qy	1314	QSTTKDSTTL-----ARI-----EMERLRIN-----EKF	1337
Db	1739	SSKSQSSASLIRSEABASVSCVHEEHTRVDESTIPSSVSGSDMPRNDINMNSKVTEEGT	1798
Qy	1338	QSOQBBIKS-----LTKERDMLK-TIKEALEVKHDQKHETRETIAKIQOESQK----	1386
Db	1799	EUSORLVRGFGAGTEIDPENEELMLNISRLOAAVEKLEIAISETSSQLEHAKVQTQELM	1858
Qy	1387	EOSLNNKEDKDNETKIVSEM-----	1410
Db	1859	RESFRQKQKQATESLKQCEELRERLHEESRAREQAVELSKAEGVIDGYADEKTLFERQIQ	1918
Qy	1411	PKDSALLRIHEIMGLSKLOESHDSEMSVAKKEDDLQRLQOEVLOSES-----DQLKENIK	1466
Db	1919	EKTDIIRLEQELLCASNRLOELEAFQOQOIEERELLSROKEMKAAGPVEQOQLLOETE	1978
Qy	1467	EIVAKHLET-----EELKVAHCCLQEETINELRVNLSEKTEISTIOKOLE	1515
Db	1979	KLMREKLEVOQAEKYRDDLOKQVKALEIDVEQVSRFIELE--QEKNTLMDLRQOQ	2035
Qy	1516	AINDKLQNK-----IQEIKYEBQOLNI-----KQISEVQENVNELKOF	1553
Db	2036	ALEKQLEKMKRFLDEQAIDREHERDFQOBIQKLEQOLKVVPFPQPISEHQ--TREVEQL	2093
Qy	1554	KEHRKADKSALESKMLLTNLOESQEEI-----QIMIKEEEMKRVQOE--	1600
Db	2094	ANHLKERTKCSSELLSKQEQORDIQERNEIEKLEFRVRELEQALLVSADTFQKVEDRK	2153
Qy	1601	-----ALQIERDOLKENTKEIVAKMKESQEKYQFLKMTAVNETQEMCE	1645
Db	2154	HFGAVEAKPELSLEVQLOAERDAIDRKEKIT-----NLEEQLEQFRE-----ELENKNEE	2204
Qy	1646	IEHLKEPFTQK-----NLNIEFTENRLTOILHENLEMSRVTKERDRLASVEETL--	1698
Db	2205	VQOLHMQLEIQKESSTTRLOEQEEN--KLPKDDMEKGLAKESDAMSTQOHHVLFQ	2260
Qy	1699	-----KVERDOLKENL-----RETITRDLEKQEBELKIVHMH--	1730
Db	2261	KFAQIIQEKEVEIDOLNEQVTKLOQOLKITTDNKBIEKNEILRDLEQIECLMSDQECV	2320
Qy	1731	KEHOETIDKLGIVSEKTEINISNN-QK-----DLHNSDALKAQDLKIQE	1775
Db	2321	KRNEEELEQNEVIEKLOELANIGQKTSMAHSLSEADSLKHQDVLVIAKLALEQO	2380
Qy	1776	LRIAH-----MHKEQOETIDKLGIVSEKTD-KLSNMOKD	1810
Db	2381	VETANEEMTFMKNVILKTNFNMNQLTQELFSLKRESEVSEKIQSPENSNNVAIDHLSKD	2440
Qy	1811	LENSNAKLOEQIOELKANEHO-----LITLKQDVNETQKQVS-----EMEQJK	1853
Db	2441	KPELEVLTB--DAKLSLENQTYFKSPENGKGSIIINLETRLLQLESTVSAKDLLETCY	2498
Qy	1854	KQIKQOS-----LTLKLEIENLNLAEQELHENLEMSKVMK-----	1889
Db	2499	KQIKQOSQOGQPETEMLOKQIVNLOKIVEEKVAAALVQIQLEAVQYAKFQDNQTISS	2558
Qy	1890	--ERDNLRRVEETLKERDOLKESLOETKARDLEIQO-----ELKTARMLSKKHETVDKLR	1944
Db	2559	EBERTNIQMLNQ--LREDELGSDISALTRLISELESQVVEHMTSLILEKEQVEIAEK--	2613
Qy	1945	EKISSEKTIQISDIQKDLQSKDELOKQKQELQOLLRVKEDEVNMHKK--INEMBQ	2001
Db	2614	-NVLEKEREKLLQKLEGNKKQKREKRSPODEVILKTTTELPHSNEESGFFNELEA	2672
Qy	2002	LKKQFEPNVLKCEMDNFOLTTKKLHESLEEIRIVAKE-----RDELRIKES	2048
Db	2673	LRAE---SVATAELASY--XEAQKLOE-ELLVKETNMTSLQKLSQVRDLHABAKK	2725
Qy	2049	LK-MERDQFIATREMIADRQNHQVQKPEKLLSDGQOHLMESLRKSKIKELLKRYSE	2107
Db	2726	LSILEKEDETFVQESKCAKCFEPLPIKLSKSIASQTDGTILKISSNQTFQI--LVKNAG-	2782

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 09:27:31 ; Search time 113.114 Seconds
(without alignments)
7428.155 Million cell updates/sec

Title: US-10-045-631B-88

Perfect score: 13329

Sequence: 1 MAEEGAVAVCVVRPLNSRE.....SOPGPHWASSGKDVPCKTQ 2663

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3615.5	27.1	2954	13	O42263 xenopus lae
2	2234.5	16.8	549	11	Q7TPX4
3	1941	14.6	699	11	Q8BWX6
4	1911	14.3	691	11	Q80YB4
5	1562.5	11.7	2244	5	Q9NCGO
6	1545.5	11.6	2013	5	Q9VKIO
7	1495	11.2	1885	5	Q869B8
8	1351	10.1	2756	10	Q9LJ60
9	1286	9.6	1931	5	Q9VKH9
10	1279.5	9.6	1931	5	Q9NCF9
11	1089	8.2	3187	11	Q63714
12	1079	8.1	1967	10	Q8GVH3
13	1077.5	8.1	1388	4	Q9NS87
14	1074	8.1	1463	5	Q9GYZ0
15	1056	7.9	1909	5	Q25893
16	1048.5	7.9	1388	13	Q91785

17	1046	7.8	3616	13	Q9W6V0
18	1036	7.8	1385	11	Q7TN17
19	1034	7.8	5767	5	Q81S25
20	1029	7.7	1385	11	Q7TSP2
21	1027	7.7	1781	5	Q86KF8
22	1004	7.5	823	10	Q9S7P3
23	999.5	7.5	2158	10	Q9LUT5
24	997	7.5	807	10	Q94HV9
25	996	7.5	2910	5	Q81BY8
26	992	7.4	2273	5	Q9U141
27	982.5	7.4	2205	5	Q9NGQ2
28	980.5	7.4	1780	4	Q9UPR5
29	971.5	7.3	888	10	Q9LQ62
30	968.5	7.3	1820	4	Q96Q89
31	964.5	7.2	7210	5	Q9V7G8
32	964.5	7.2	9270	5	Q8MLD9
33	958	7.2	1695	3	Q86ZA2
34	957.5	7.2	1127	4	Q86TN3
35	956.5	7.2	2867	5	Q9N2M3
36	955.5	7.2	1231	11	Q80YP3
37	951	7.1	1844	5	Q81S00
38	946	7.1	1939	5	Q25662
39	938.5	7.0	1596	5	Q81J44
40	937	7.0	2473	11	Q9QZ84
41	928.5	7.0	1968	3	Q8X0C5
42	927.5	7.0	1831	10	Q7XN11
43	917	6.9	2770	5	Q7YUE9
44	916	6.9	3256	5	Q7YWE8
45	915.5	6.9	1401	4	Q86VH2

ALIGNMENTS

RESULT 1

O42263 ID O42263 PRELIMINARY; PRT; 2954 AA.
AC O42263;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DE 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Kinesin-related protein.
GN XCENP-E.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP MEDLINE=98028574; PubMed=9363944;
RA Wood K.W., Sakowicz R., Goldstein L.S., Cleveland D.W.;
RT "CENP-E is a plus end-directed kinetochore motor required for
metaphase chromosome alignment.";
RL Cell 91:357-366 (1997).
DR EMBL; AF027728; AAC60300.1; --
DR PIR; T14156; T14156.
DR HSSP; P33176; 1BG2.
DR GO; GO:0005871; C:kinesin complex; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR GO; GO:0007017; P:microtubule-based process; IEA.
DR InterPro; IPR001752; Kinesin_motor.
DR Pfam; PF00225; Kinesin; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
DR PROSITE; PS50067; KINESIN MOTOR DOMAIN2; 1.
SQ SEQUENCE 2954 AA; 339964 MW; 439804ED0E592679 CRC64;

Query Match 27.1%; Score 3615.5; DB 13; Length 2954;
Best Local Similarity 32.0%; Pred. No. 5.9e-100;
Matches 996; Conservative 572; Mismatches 932; Indels 615; Gaps 86;

1 MABEGAVAVCVVRPLNSREESLGETAQVYWKTDNNVIYQVDSKSFNFDRVPHGNBTX 60
 1 MSGGDVAVCVVRPLRIQREQ--GDQANLQWKAAGNTISQVDTGKSFNFDRVFNHSHTS 58
 61 NVYEEAALPIDSALQSYNGTIFAYGQTSGKTYTMMGSEHDHGVTPRAIHDIFQKIKKF 120
 59 QIYQEIAPVPIRSALQSYNGTIFAYGQTSSGKTYTMMGTPNSLGLIIPQAIQEVFKIIOEI 118
 121 PDREFLLRVSMYELNYETITDILCGTCMKPLIREDVNVNNAVYADITEEVYTSMAK 180
 119 PNREFLLRVSMYELNYETITDILCGTCMKPLIREDVNVNNAVYADITEEVYTSMAK 178
 181 WITKGEKSRHYGETMMORSRSHITFRMILESEKGEPS--NCEGSVKVSHNLVDLA 237
 179 WIKKGNRYGETMMORSRSHITFRMILESEKGEPS--NCEGSVKVSHNLVDLA 238
 238 GSRAAQCTGAAGVRLKBCGNINSLFILGVQVVKLSGQVGGFNYNRDSKLTTRILQNSLG 297
 239 GSRAAQCTGAAGVRLKBCGNINSLFILGVQVVKLSGQVGGFNYNRDSKLTTRILQNSLG 298
 298 GNPTRITICTTPVSFDETLTALQFASTAKYMKNTPVYNEVSTDEALLKRYKREIMDLKK 357
 299 GNAKTIVICTTPVSFDETLTALQFASTAKYMKNTPVYNEVSTDEALLKRYKREIMDLKK 358
 358 QLE--EVSLETRAQAMEKQOALQLEKDLQVQNEKIENLRLMVTSSSLTLQOELKA 415
 359 QLENLESSETKAQAMEKEHTQALLAEIKQLHKEDEORIHWTIIVASSQES--QDQVR 417
 416 KRKRRTVCLGKINKMKNNSYADQFN--IPTWITTKHKLINLREIDESVCSRDV 471
 418 KRKRRTVWAPGKI QNSLHASGVDFDMLSRPLPGNFSKAKFSDMPSPFEIDDSVCTEFS 477
 472 FSNITDLTSL--ELENPAKTLNQE----- 494
 478 FDDALSMWDSNGIDAEMNLASKVTHREKTSLSHQSMIDFGQISDSVQFHDSSKENQLYLP 537
 495 -----NTESE 499
 538 KDSGDMACRKAASPEKEITSLQOQLOQSEKEKEIKVQSFEKIAEBLEBQLSVKAQNLV 597
 500 LNS----- 502
 598 TNSREHSINAEVQTDVEKEVVRKXMSVLGSDGYNASDLQDSSVDGKRLLSSSHDECIEH 657
 503 -----LRA 505
 658 RMLEQKIVDLEEFTEINLKKSENDKQSSQDFMESIQOLCEAIMAEKANALEALMRD 717
 506 DYDNLVDLYEQLRTEKEEMELKKEKNDLDEFEALERKTKKQDOEMQLIHEISNLKMLVKH 565
 718 NFDNILLNEVTLKREIADLRSLENQENETNEFEILEKETQKEHEAQLIHEIGSLKLVEN 777
 566 REVYNQDLENESSLKVELLREKEDQIKLQEVDSQKLENIKMDLSYSL-----ESIEDPK 621
 778 AEMYQNLEEDLETTKLKEQEIQLAELRKADNLQKVRNFDLSVSGMDSEKICEEILF 837
 622 QMKQTLFPAFTVALDARESALRSLENLEKMKELATYKQWENDIOLYOSQLEAK-- 679
 838 QLKQSLSDAEAVTRAQKESFURSENLEKMKEDTSWNWYQKKAASLFPKQLETEKS 897
 680 --KMQVLEKELQSAFNEITKLTSLIDGKVPKDLCLCNLEEGKITDLOKEANKEVEENE 737
 898 NYKKMEADLOKELQSAFNEIYNGLAGVPRDLSRVELEKVKVSEFSQKLEALEEKN 957
 738 ALREEVILLSELKSPSEVERURKEIQDKSELHIITSEKDLFSEVHVHKESEVQGLLAE 797
 958 ALENEVTCISEYKFLPNEVECLKNQTSKASEBIMLLKQGEHSASIIISKQEIIMQOSEQ 1017
 798 IGKTKDLDLATTOSNYKSTDOEQNFKTLHMDFEQYKMWLEENERNNOEIVNLSKQAKRF 857
 1018 ILQLTDEVTHTOSKVQOQTEQYLEMKKMDDDLFEKY-----IRNKSEADLLREMN 1070

858 DSSLGALKTSLSYKTOELOKREVOERLNEBOLKEOLEN--RDSPLQTVREKTELITE 915
 1071 KGTMESVYVYADTKHELEETITNDKEQLLHEKXFFQAMQTIPTIFPLSDSLSPSKLVEG 1130
 916 KLOQTLE--EVKTLTQEKDDLOKLOESLQIERDOLKSDIHDVTNMMIDTQOELRUALES 973
 1131 NSQDPIEINDYHNLIALATERNNIMVCLETERNSLKEV-----IDLNTOLQ-SLQAO 1182
 974 KOHQETINTLKSISE--EVSRLNHEENTGETKDEFQOKMVGIDKQ--DLE-AKNTQTL 1029
 1183 SIEKSDLOPKQDLGEVKKLLEMBELKGLHTD---SOLSTIEKQLEMLEVTEKLOTL 1238
 1030 TADVKONEIIEQOKRIFSLIQEKNELOQMESVIAEKEOLKTDLKENIEMTQENQEBRL 1089
 1239 QEMKNITI-----ERNELQTNFEDLKAHDSLKQDSENIEQSIETQDELRA 1286
 1090 LGDELKQKQBIQAQEKNAHAIK--REGELSRCTDLAEVVEKKEKQSOLOKQOQLLVQ 1148
 1287 AQELREKQQLVDSFRQQLDSCSVGISSPNHDAVANQEKVSLGEVNSLQSEM-----LRG 1341
 1149 EMSEMOKK-----INEIENKVELKNKELTLEHMETEELAEKQKLENNYEVKSTTKERKV 1204
 1342 ERDELQTSCKALVSELELLRAHVK-----VEGENLEITTKLNGLEKEILGKSESEV 1394
 1205 LKELQKSFETERDLRGY-----IRIEATGLQTKBELKIAHILKHEQBTID 1252
 1395 LKSMLENDKDNKKLKEQAEYSKENQFSLSEVFSQKLVDBIEVLKQAKAAERLE 1454
 1253 ELRSRSEKTAQIINTQDLE--KSHTKLOEPIVLEHEQELLPNVKVSEFQETNNELELL 1311
 1455 IKORDYFE--LVQPTANTNLVEKLETPLOAD-----HEED-----SIDRSEEM--EIKVL 1501
 1312 TQSTTKDSTTLARIEMERLRLNEKFBQSEIEKSLTKERDNKATIEALVKHDKHEH 1371
 1502 GEK--LERNQVLLERLOEKLELSNKLRILOKEMETSVLLDDLOKQLESLENNILKEN 1560
 1372 IRETL-----AKIQSSQSKOEOLNWKEDNTTKIVSMEQFKPKDSALLRIEIML 1424
 1561 IDTTLKHSHTQAOLOKTKQOELQALAKNLAIASDNCPIQKE---TSADCVHPLEEKIL 1617
 1425 GLSKRLQESHDMSVAKEDDLOLQEVLSQESDQ-----KENIKEIVAKHELETE 1477
 1618 LLEELHQKTVNEQEKLLHEKNELEQAQVELKCEVHELMKSMIESKSSLESQHEKHDTEQ 1677
 1478 ELKVAHCCLEKQEBETINELRVNLSEK-----TEISTIQOLEANDKLQKQIEYKBE 1533
 1678 QL-----LALKQOMQVVTQEKKELOQTHEHLTARVDHLKENIE--LGINFNRAQKTYTEQ 1732
 1534 QL--NIKOISVQENVN--ELKQFKEHRAKADSAQSI---ESKMLELTNRLQESQEEIQI 1587
 1733 CLLENKELESQHLQCEIEELMKSJKDKESALETLKESBQKVINL-----NQEMEM 1785
 1588 MIKEKEEMRVQEAQIERDQKENTKEIVAKMESQB--KEYQFLKMTAVNETQKCE 1645
 1786 VMLEEMELKNSQRTVIAERDQLODDLRSEVMSIETODDLRAQALQOQKQKVELTSQ 1845
 1646 IEHLKEQFETQKLNLEIETENIRLTQILHENLEEMRSVTKERDRLRSEETLKVVERDQ 1705
 1846 ISVLOEKISL-----LEN-----OMLY--NVATVKETLSERDOLNOSKOHFSEIETL 1891
 1706 KENLETTITRILEKQELKIVHMLKHEQETID---KLRGIVSEKTNISNNQKDLHNS 1762
 1892 SUSLKEK--EPALQEAQEK-----ADAARKTIDITEKISNIEEQLQATNLKETL--YBR 1944
 1763 DALKAQDLKIQEBELIAHMLKQOQETIDKLRGIVSEKTDKLSNNQKOLENSAKLOEKI 1822
 1945 ESL-----IQCKEQALNTEHLRETILKSKDALGKQGEDEAANKVIALTEKMSLEBOI 2000
 1823 QE-----LKAHEQHLTLKQVNETQKYSMEQLAKQIKQOSILTSLK-----ETB 1869
 2001 NENVVTLKEGEKETFYLQRPFSKQOSSQMBELRESLTKOLOEAEKEISEATNEIK 2060
 1870 NL-----NLAQELHENLEBMSVMEKRONLRVEETLKLERDQKLESQETKARDLEIOQ 1924

```

Db      2061  NITAKISLESEILQALNPAVERENLRHSKQOLVSELEQLSLTL---KSRDHAFQAQ 2117
QY      1925  ELKTARMLSKGHETQVLDREKISIKTIQISDIQKLDK---SKDELQKIQELQ----- 1976
Db      2118  -----SKREK---DEAVNKIASIAEIKILTKEDEDFRDSKESLQEQSSHJUSEELCT 2166
QY      1977  -KKEIQLLR-VKEDVMS-HKKINMEQLKQFEPNYLCKEMDNFQTKKLHESLEIR 2033
Db      2167  YKTELQMLQKQEDINNKLAEKVKVEDEL-----LQHLSSLKQEQDQIQ 2210
QY      2034  IVAKERDELRIKESLKE-RDQFATLREMIARDQNHQVKEPKLLSDCQOHLMESLR 2092
Db      2211  MELR-NEKRLNYELCEKMDIMEKISVLRML-----ONEPOQE---EDDVAERMDILE 2259
QY      2093  EKCRIKELRLKRYSEM-DDHYECLNRLSLDLKEIEFHR----- 2130
Db      2260  SRNOEIQELMEKISAVYSEQHTLLSSLSSELQKETEAKHCKMNIKESLSLSTLSRSPGSL 2319
QY      2131  -----IMKKIKYVLVYVYKIKERQHECINKFEMDPIDEVEKQKELIKIQHL 2177
Db      2320  QTEHVKLNTQTLNLKFKVY--YRTAAVKEDHSLINDYKDLAAEQKRHDLRLQLQCL 2377
QY      2178  QOD-----CDVPSRELRLKLNQNDL-----HIEETLKDFSESE--FPSIKTEFOQVLS 2225
Db      2378  EQHGRKWSDSASBELKCEIEFNEELLFKKANIQSVDQDFSEVQVFLNQVSGTLOBELE 2437
QY      2226  NRKEMTOFLEWLNTRPDIEKLNKIQKENDRICQVNNFNRRILAIWNSTEFERSAT 2285
Db      2438  HKKGFQMWLEEFGLHVDAAKLSSEGQENRRITASTITQLTKLKVQSKI---QREIT 2494
QY      2286  I-SKEWEQDLKSLKEKNEKLFKNYQTLKYTSLAGAQVN-----PTQDNKNPHVTSRA 2337
Db      2495  VYLNQPEAKLQEKQKQKELMRMEHHPGASVMEENARLLGILKTQDE-----SKK 2548
QY      2338  TQLTTEKIRELENSLHEAKESAMHESKLIKQKQLEVTNDIITAKL---QAKVHESNKCL 2394
Db      2549  LQ---SRIKMLENELNVQDDAMHKEKVAILODKL-LSRNAEAEELNMQVKLTQKQDNL 2604
QY      2395  EKTETIQVLQDKVALGAKPYKEIEIDLKMLKVIDLEKMKNAKBPKEISATKATVEYQ 2454
Db      2605  QAAMKEIENLQWAKGAVPYKEIEDNLKTKVVKIEMEKIKYSKATQDIAVLKSCLEDK 2664
QY      2455  KEVIRLLRENLSOQAQTSVISEHTDOPSNKPLTCGGSGIVQNTKALIKSEHRL 2514
Db      2665  EGLRLKEELRRQAQNDTTCVPKDYQKASTFPVTCGGSGIVQSTAMVLQSEKAL 2724
QY      2515  EKEISKLAQONQOL-----IKQKELLNNQHLNNEVKTWKERTILKRAHKQV 2562
Db      2725  ERELHYKKVYHLSTRMSSSEDRKTKAKSDAHSHTGSSHRGSPHKTETYR---HGFV 2781
QY      2563  TCE-----NSPK-----SPKVTGT-----ASK 2579
Db      2782  TPERSEMPSLHLGSPKSESSTKRVVSPNRISYSLVMSPKTQWHKHLISPSKVGLHK 2841
QY      2580  KKQITP-----SOCKERNLQDPVPKSPKSCFF 2607
Db      2842  KRALSPNRSEMPTHVSPGKTGLHKNLTESTFLFDNLSPCKQKQVQENL--NSPKGLF 2899
QY      2608  DSRKSLSPHPVRVFDNSSLGLCPQVQAGAESVDSQPGWPHASSGKDVPECKT 2662
Db      2900  DVKSKSMFY-CPSQPFQNSKLQDFSELNTAESNKSQAENWVYAKKETAPECKT 2953

RESULT 2
Q7TPX4
ID Q7TPX4
AC Q7TPX4;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DS Hypothetical protein (Fragment).
OS Mus musculus (Mouse).

```

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Egg;
RX MEDLIN=23389257; PubMed=12477932;
RA Strausberg R.B., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Brownstein M.J., Sodes M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Raha S.S., Loquellano P.J., Toshiyuki S., Carninci P., Prange C.,
RA Bobak S.A., McEwan P.J., McKernan K.J., Abramson R.D., Mullahy S.J.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gunaratne P.H.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Schmutz J., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
[2]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Egg;
RA Strausberg R.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC052843; AAH52843.1; -.
KW Hypothetical protein.
FT NON TER 549
SQ SEQUENCE 549 AA; 62899 MW; B707C97DC9FB38D3 CRC64;

```

```

Query Match 16.8%; Score 2234.5; DB 11; Length 549;
Best Local Similarity 79.2%; Pred. No. 1.5e-59;
Matches 437; Conservative 46; Mismatches 60; Indels 9; Gaps 2;

QY 1 MAEGAVAVCVVRPLNRSRESIGETAQVYKTDNNVYQVDSGSKSFNDRVPHGNETTK 60
Db 1 MAEEASVAVCVVRPLNRSREELGEATHYKTDKNAYIQSDGSKSFQDRVDSNETTK 60
QY 61 NYVEEIAPIIDSAIQYNGITFAIGQTASGKYTMGSEDLGVIPRAIHDIPOKIKKF 120
Db 61 NYVEEIAVPIISSAIQYNGITFAIGQTASGKYTMGSEDLGVIPRAIHDIPOKIKKF 120
QY 121 PDRELLIRVSYMEIYNETITDLCQTKMKPLIREDVYNNVYADLTTEVVYVTEEMALK 180
Db 121 PEREFLIRVSYMEIYNETITDLCNAQKMKPLIREDVYNNVYADLTTEVVYVTEEMALK 180
QY 181 WITKESRHRGKETKONORSRSHITFMILRESKPEGPNCEGSKVSHNLNVLDSAGE 240
Db 181 WATGEKRNHRGITKONORSRSHITFMILRESKPEGPNCEGSKVSHNLNVLDSAGE 240
QY 241 RAAQTGAAGVRLKEGCMINRSLFTLGOVTKLSDQVGGFNYRDSKLTILQNSLGNP 300
Db 241 RAAQTGAAGVRLKEGCFINRNLFTLGOVTKLSDQVGGFNYRDSKLTILQNSLGNP 300
QY 301 KTRIICITVPSFDETTALQFASATKYMKNTPYVNEVSTDEALLKRYRKEIMOLKKOLE 360
Db 301 KTRIICITTPASLDETTLTQFASATKYMKNTPYVNEVSTDEALLKRYRKEIMOLKKOLE 360
QY 361 EVSLERTRAQAMEKDLQALLKEKDLQKQVNEKIENITRMVLTSSSITLQOELKAKRRR 420
Db 361 EVNTKTRAQAMEKDLQALLKEKDLQKQVNEKIENITRMVLTSSSITLQOELKAKRRR 420
QY 421 VTWCLGKINKMKNVADQFNIPNTITTKTKLSINLIREI-----DESVCSDVFSN 474
Db 421 VTWCYG---KMKDSNKEKEKVFVTSITTRKRTSVTSLRENSLMKFGESAASEFEMINN 477

```

262 EVEKQNELQIQSLSQTYRIPARDLQ-IKLSQMDLHI EEMLKDFSENDFITIKTEVQQ 320

2223 VLSNRKEMTQFTLEWLNTRFDIEKLNKGOKENDRICOVNPNFNRIIATMINESTEFEE 2282

321 VLNNRKEITEFLGKWLNTLFDTENLKSTIQENKSGILGVNPNYHSRIITAMINESTEFEE 380

2283 SATISKEMEQDLKSLKEKNEKLFKNVQTLKTS---LASGAQVNPPTQDNKNPHVTSRATQ 2339

381 SATRSKDLQOYLKSLKETTEQLSEVQTLTASQSVVHLHPTVOPSTROSERPQAASGAQ 440

2340 LTTEKIRELENSLHEAKESAMHESKIIKQKELEVTNDIIAKLOAKVHESNCKLEKTKE 2399

441 LTS-----443

2400 TIQVLQKVALGAKPYKEETEDLKMVLKIDLEKKNKAKEFEKEISATKATVEYQKEVIR 2459

444 ----KNKIALGAVPYKEETEDLKMQLVKSDDLKATKATKAFKYLKSLKATVEHQEMIR 498

2460 LLRENLRSSQAQDTSVISEHTDPPQSPNPLTCGGSGIVQNTYKALILKSEHRLKEIS 2519

499 LLRENLRGHOAQDTSWISQ-DSQLLSKPLTCGGSGIVQNTYKALILKSEYKRMGSEIS 557

2520 KLKQONEQLIKQKNELLSNQHLISNEVKTWKETLREAHKQVTCENSPKSPKVTGTASK 2579

558 KLKQONEQLKQNNQLLSNLSNEVKTWKETLREAHKQVTCENSPKSPKVTGTASK 617

2580 KQITPQCKERNLODPVKESPKSCFFDSRSKSLSPHPVRYFDNSSLGLCPFEVQVAGA 2639

618 RRONTTSQCAQNLQDPVKESPKSWFFDNRSKSLPAPHPIRYFDNSSLGLCPEDD--V 675

2640 ESDVDSQPGMHASGKDVPECKTQ 2663

676 ENVEPKTDLQASLEKDVQCKTQ 699

RESULT 4

Q80YB4 PRELIMINARY; PRT; 691 AA.

ID Q80YB4

AC Q80YB4; 01-JUN-2003 (TRENBLrel. 24, Created)

DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)

DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)

DE Similar to centromere protein E, 312kDa (Fragment).

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NCBI_TaxID=10090;

11

SEQUENCE FROM N.A.

RP TISSUE=Olfactory epithelium;

RC Strausberg R.;

RA Submitted (MAR-2003) to the EMBL/GenBank/DBSJ databases.

RL EMBL; BC049989; AAH49989.1; --

DR GO; GO:0005215; F:transporter activity; IEA.

DR GO; GO:0006810; P:transport; IEA.

DR InterPro; IPR000566; Lipocln_cytFABP.

DR PROSITE; PS00213; LIPOCALIN; 1.

FT NON_TER 1

SQ SEQUENCE 691 AA; 80186 MW; 1D12F180EEC17078 CRC64;

Query Match 14.3%; Score 1911; DB 11; Length 691;

Best Local Similarity 46.0%; Pred. No. 7.9e-50;

Matches 421; Conservative 110; Mismatches 119; Indels 266; Gaps 10;

QY 1789 TIDKLGRIVSEKTDKLSNNKQDLENSNAKLOEKIQELKANEHQLITLKDVNETQKVSSE 1848

DB 1 TIDRLMETNSEKTEEELSNMKNLENNVNMKLOEKVQELKTSERQVRLKADASEA----- 54

QY 1849 MEQLKQIQKQDSLTLSKLEIENLTAQELHENLEEMKSVMKERDNLRRVEETLKLRDQL 1908

DB 55 ----KKELKQGLTSLKLEIENLTAQELHENLEEMKSVMKERDNLRRVEETLKLRDQL 110

QY 1909 KESIQETKANDLEIQELKTAARMLSKEHKEVTDKLRKIKSEKTIQISDIQKLDKSKDEL 1968

475 TLDLTSELEWNPATKLINOENISELSNLRADYDNLVLDYEQLRTEKEEMELKLEKNDL 534

478 TLESIAEVESSATLLISEENVESELNLAQYNDLVLDYEQLRRENEDELKLEKNEKL 537

535 DEFEALERKTKK 546

538 BEFELLEQRKKK 549

RESULT 3

Q8BWK6 PRELIMINARY; PRT; 699 AA.

ID Q8BWK6

AC Q8BWK6; 01-MAR-2003 (TRENBLrel. 23, Created)

DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)

DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)

DE Hypothetical lipocalin-related protein and Bos/Can/Equ allergen containing protein.

DE Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NCBI_TaxID=10090;

11

SEQUENCE FROM N.A.

RP STRAIN=C57BL/6J; TISSUE=Spinal cord;

RC MEDLINE=22354683; PubMed=12466851;

RA THE FANTOM Consortium,

RT "Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs."

RT Nature 420:563-573(2002).

RL EMBL; AK049676; BAC33868.1; --

DR GO; GO:0005215; F:transporter activity; IEA.

DR GO; GO:0006810; P:transport; IEA.

DR InterPro; IPR000566; Lipocln_cytFABP.

DR PROSITE; PS00213; LIPOCALIN; 1.

KW Hypothetical protein.

SQ SEQUENCE 699 AA; 81197 MW; 7392ECC11656C147 CRC64;

Query Match 14.6%; Score 1941; DB 11; Length 699;

Best Local Similarity 46.2%; Pred. No. 1e-50;

Matches 427; Conservative 111; Mismatches 120; Indels 266; Gaps 10;

QY 1781 MHLKEQETIDKLGRIVSEKTDKLSNNKQDLENSNAKLOEKIQELKANEHQLITLKQVNV 1840

DB 1 MNLKEHQETIDRLMETNSEKTEEELSNMKNLENNVNMKLOEKVQELKTSERQVRLKADAS 60

1841 ETQKVSMEQKQIQKQDSLTLSKLEIENLTAQELHENLEEMKSVMKERDNLRRVEET 1900

DB 61 EA-----KKELKQGLTSLKLEIENLTAQELHENLEEMKSVMKERDNLRRVEET 110

1901 LKLRDQLKESIQETKARDLSIQELKTAARMLSKEHKEVTDKLRKIKSEKTIQISDIQKD 1960

DB 111 LRWERDQLKDN-----121

1961 LDKSKDELQKIQELKQELKQLLRVKEDVNMVSHKKNEMEQKXQFEPNLYCKEMDNFQ 2020

DB 122 -----121

2021 LTKKLHESLEIRIVAKERDELRRIKESIKMERDOFIATIRMIARDONHQ--VKPEKR 2078

DB 122 -----LRAMUKAHQNHETWKCKG 142

2079 LLSGQOHLMESIREKCSRIKELLKRYSEMDHYECLNRLSLDLEKE----- 2125

DB 143 LKACAG-EYCTGLREKCFRIEKLKRYSEMDHYECLNRLSLDLEKEKTSVTVRT 201

2126 -----IEFHRIKMLKXVLSVVKIKKEQHECINKFEMDID 2162

DB 202 KLSLPHQTQKEMELKATANQRCLEPHRALKELKVLSSLIARIKEQHECINKFEMAFIQ 261

2163 EVEKQKELLIKIHLQODCDVPSRELRLDKLNQMDLHLEELKDFSEFPSPKTEFQQ 2222

DB 111 KDN----- 113

QY 1969 QXKIQLOKKELOLLRVEDVNVSHKKNEMEQKKQFEPNYLCKCEMDFNOLTKKLHES 2028

DB 114 ----- 113

QY 2029 LBEIRIVAKERDELRIKESLKMEDQFIATLREMIARDRONHO--VKPEKRLSDGQOH 2086

DB 114 ----- 113

QY 2087 LMSIREKCSRIKELLKRYSEMDDHYECNLRLSLDLKE----- 2125

DB 142 CTGRLREKCFRIKELKRYSEMANDYECNLKVSJDLERETKQKLSVTVRTKLSLPHQTQ 201

QY 2126 ----- 2170

DB 202 TKEMKLLTANORCSLEFHRAUKLYVLSSTARIKEQEHESINKREMAFIOVEKQNEL 261

QY 2171 LIKIQHLOQDCDVPRELRDLKLNQMDLHIEILKDFSEFPSSIKTEFOQVLNKRKM 2230

DB 262 QIQISLSQTYRIPARDLQ-IKLSQEMDLHIEMLKDFSENDFLTIKTEVQVQLNKRKEI 320

QY 2231 TOFLEWLNTRDIEKLKNGIOKENDRICQVNNFFNNRIIAIMNESTEFERSATISKEW 2290

DB 321 TEFLGKWLNTLFDTENLAKSTIQENKSGLVNPFYHSRITAMINESTEFERSATRSKOL 380

QY 2291 EQLKSLKKEKNEKLFKNYOTLKT-----LASGAVNPTQDNKNPHVTSPATQLTTEKIRE 2347

DB 381 DOYLSKLTETTELSEVYQTLTASQSVVHLHTVPTQSTRDSRPOAASGAELTS----- 435

QY 2348 LENSILHAKESAMHKEKIIKMQKELEVTNDIIAKLOAKVHESNCKLEKTKETIQVLODK 2407

DB 436 ----- 438

QY 2408 VALGAKPKKEIEDLKMVLKIDLEKKNKAKEFEKESATKATVBYOKEVIRLLRENLR 2467

DB 439 IALGAVPYKEBEDLKMQLVKSLDEKKAKEFEKESATKATVBYOKEVIRLLRENLR 498

QY 2468 SQOQDTSVSEHTDPOSNKPLTCGGSGVQNTKALILKSEHRLKESIKLKQNEQ 2527

DB 499 HQOQDTSVSEHTDPOSNKPLTCGGSGVQNTKALILKSEHRLKESIKLKQNEQ 557

QY 2528 LIKQKELLNNHLSNWKVKTBLKRAHKQVTCENSPKSPKVTGTASKKQITPSQ 2587

DB 558 LRKQNLQSLNSQSLSEVKEWETLARDSYRETTENSPKSPKVTGTASKKQITPSQ 617

QY 2588 CKERNLODPVKESPKSCFFDSRSKSLSPHPVRYFDNSSLGLCPVQVAGAESYDSQPG 2647

DB 618 CRAQNLQDPVKESPKSCFFDSRSKSLSPHPVRYFDNSSLGLCPVQVAGAESYDSQPG 2647

QY 2648 PWHASSGKQVPECKTQ 2663

DB 676 LCQASLEKQVSOCKTQ 691

RESULT 5

Q9NCGO PRELIMINARY; PRT; 2244 AA.

AC Q9NCGO;

DT 01-OCT-2000 (Tremblrel. 15, Created)

DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)

DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)

DB Kinesin-like kinetochore motor protein CENP-meta.

DB CMT OR CG6392.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

NCBI_TaxID=7227;

RP SEQUENCE FROM N.A.

RX MEDLINE=20351410; PubMed=10893249;

RA Yucel J.K., Marszalek J.D., McIntosh J.R., Goldstein L.S.B.,
RA Cleveland D.W., Philp A.V.;
RT "CENP-meta, an Essential Kinetochore Kinesin Required for the
RT Maintenance of Metaphase Chromosome Alignment in Drosophila";
RL J. Cell Biol. 150:1-12(2000).
DR EMBL; AF220353; AAF32355.1; --
DR HSSP; P33176; 1BG2.
DR FlyBase; FBgn0040232; cmet.
DR GO; GO:0005699; C:kinetochore; IDA.
DR GO; GO:0007080; P:mitotic metaphase plate congression; IMP.
DR InterPro; IPR001752; kinesin_motor.
DR Pfam; PF00225; kinesin; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.
DR PROSITE; PS00667; KINESIN MOTOR DOMAIN2; 1.
SQ SEQUENCE 2244 AA; 257992 MW; FA6AA3B2A541ADE0 CRC64;

Query Match 11.7%; Score 1562.5; DB 5; Length 2244;
Best Local Similarity 25.0%; Pred. No. 5.4e-39;
Matches 645; Conservative 463; Mismatches 847; Indels 621; Gaps 106;

QY 6 AVAVCVRVPLNSREESLGETAQVYWKI-DNNVIYQVDG-SKSFNDFRVHGNETTKVY 63
DB 8 SIQVCIKVRPCEPGLTSL-----WQKEGSIHLADSHAEPPYVDYVDEGASQNEVF 60

QY 64 BEIAAPIIDSAIQVNGTIFAYGOTASGKTYTMMGSEDLGVIPRAIHDFOKIKKFPDR 123
DB 61 DRMARHIVHACMQGFGTIFAYGOTSSGKTYTMMGDEQNGVWLAAKEIFQOISSETER 120

QY 124 EFLLRVSYMEIYNEITDILCGTQKMKPLIREDVNRVYVADLTVEVYVYSEM-ALKWI 182
DB 121 DFLRVGVYIEYNEKYDILL--NKKNDLKHESGNGVNVN--CEECIITSEVLLRL 176

QY 183 TKGEKSRHYGTMMNORSRSHITFRMILEGREKGEPSNCEGSKVSHNLNVLDSGSA 242
DB 177 CLGKERTVGTNNMERSRSHAFKIIIEGR-KSDHSD-DDAVTQSVNLNVLDSGSA 234

QY 243 AQTGAAGVRLKEGNCINRSLFILGQVKKLSDGVGGFINTYRDSKLTILQNSLGNPKT 302
DB 235 DOTGARGARLKEGGHINKSLFLSNVKSLENADNRFTNYRDSKLTILQNSLGNPKT 294

QY 303 RIICITIPVSFDELITLQFASAKYMKQNTVYNEVSTDEALLKRYRKEIMDLKQLEEV 362
DB 295 SIICITIPVSIMEESQSTLSPATRAKIRIKPQVNEVSDATWMLKRLEREIKVLKDLAE- 353

QY 363 SLETRAQAMEKDLQALKEEDLLQKVQNEKIENTM-----LVTSSTSLTQOELKA 415
DB 354 -----EER--KNENQKVHLEHRIKHMKEIKICGHSL-----DKG 388

QY 416 KRKRRTWC-----LGKINKMKNYADQF---NIPTNITTKTKL-----SI 455
DB 389 QOKRRRTWCPTASGSHLEAETGTEDRIDQFPKVSHPKPFVFTSNAGKRWNPKIT 448

QY 456 NLAREIDESVCSDESNTLDTLSEIENWPATKLQNEIESINLADYDNLVDYE 515
DB 449 NILGSLD--IGTES-----NSISE-EFLPAECI-----DFGSPRPDLKPLMTIR 490

QY 516 QLRTEKEEMELKKEKNDLDEFEALERKTKKQOEQLIHEISNLKLVKHEVYNQOLEN 575
DB 491 QL-----PDLPLTPKGPL-----TTDKIKKE-----IQDLQMTSLKHFVEEVEEQ 534

QY 576 ELSSKVELLREKEDQIKKQLEYIDSKLENIKMDLSYSLESIEDPKQMKQTLFDAETVAL 635
DB 535 -----LKEKLAETVAQRDNLEQESL-----AEKERYDALEKEVTS 570

QY 636 DAKRESAPLSENLKELKKEKELATTYKQEMNDIQL-----YQSQLEAKK---KMQVD--- 685
DB 571 RADNEAA--NSKISELEKSLTKQTMRIEMEVNQAVGLEFEFEFAKSKSLKRDLLS 628

QY 686 --LEKE-----LOSAFNEITK--LTSIDG-----KVPKDLICNL--ELEKIDT 724
DB 629 ALLEKSTIESLQSLNLTDRVLRNKSHEGMLSIAPQEDVAGDSICNKEELEKLIAD 688

1702 RDOLKENRETITRDLEKQBELKIVHMLKEHOETIDKLRGIVSEKTNBISMNQDLEHS 1761
1615 ANELSNLQKV-----MSLHELIDSQGI-SSRDVEINELRELKAA 1657
1762 NDA---LKAQDLKIOBELRIAHMLKEQOETIDK---LRGIVSEKTKLSNMQ---KD-L 1811
1658 MDAKATASAEQMTLVTLQKDVBERMANOAEKFTREANLKGSIINELLKINSMOETKDL 1717
1812 ENSNAKLOKIOB-----LKANEHOLITIKDVNETOKKYSEMEQL---KQIKDQS 1860
1718 ESGNEBELQELRNSQNLNMLDEESKMCISLKEKLVKLEDAKTSLQQLADNKSEIYQRH 1777
1861 LTLKLEIE-NLNLAQELHENLEEMKSVKMRNLRRVBEETLKLERDQKESQETKARD 1919
1778 TELTK-EVELGRNRIKELTKCEELCSLENSQIR-----LDI-ETKEQLKTLLENN 1829
1920 LEIQOELKTARMLKSEKHEKTVDKLRKISEKTIQISDIQKDLKSDKDELOKKEOELOKKE 1979
1830 LGWQQ-----KQVDEVTRECEKLRPFDMQSKVQESKVOEL----- 1864
1980 LQLLRVKEDVNNSHKKINEMEQKKQFEPNLYLCKEWMDFQTLKLLHESLEEIRIVAKER 2039
1865 -----ISECELRSTL-----KSEASQSEK-----ESMD---RTTSSIL 1897
2040 DELRRIKESL-----KMERDOFIATLREMIARDRONHQVKPEKRLSDGQOHLMESL 2091
1898 EDRNLEELKLSANDIVAKLETE---IAALRPKSLDR---NPVPRKSITFE----- 1943
2092 REKCSRIKELLKRYSEMDDHYECLNRLSLDLKEIEFHRIMKKLY-----VLS 2140
1944 -----SEIRKN-----RRISVHDERRSQVWVDFEFGIMTDPVGMSSCVFP 1995
2141 YVTKIE---EHECINKFEMDFIDEVKEQKELLIKIQLHQDCDVPSERLDLKLNQM 2197
1986 LVFPLNESFRQNC-----NCAELNSKLDQCORELFI---RESQVTLKXEL 2030
2198 DLHIEETLKDFSESEPPSIKTEFQOVLNKRKMTQFLEEWLNTDFIEKLNQIOKXENDR 2257
2031 DHH---PLKO-----ENALTKRVIEEQDKAKVEQKRLAMKLODLNAR 2070
2258 ICQVNNFFNRIIATIMNESTEFEB--RSATIS---KEWQDLKSLKEK-NEKLFPNYQTLK 2312
2071 INDLTT-----ASAKEPESNQMAQAKPATVAAQTQTESDLETILEKTNVYEEAVRMLR 2125
2313 TSLASQAVNPTTQDNKPNVTSRATOLTTIEKIRELENSLHEAKESAMHESKIIK 2368
2126 YRYNLIQELKEKLRQNSD-TSNITLSAGQTSALKVAI-----SNRYSSSYK 2175

725 LOKELNKEVEENALREEVILLSELKSLPSEVERLKRKEIQDKSEELHIITSEKDLFSEV 784
689 LESKNS-----CECDQLREIVSRVDKLESVESAFNLASSGI 726
785 VHKESSVOGLLEBEIGTKDKDLATQSNYKSTQBFQ-----NFKTLMDFE---OKYKWL 837
727 IQKATDCERLSKELSTSNQAFQQLQRYDALDOQMAQOAGIITLNEHEHVEOIEYOKLQ 786
838 ENERM-----NOEIVNLSKEAOKFSSIGALAKTSLSYKTOELQEXTREVOERLNE 889
787 EYEQELSRARSASSAEFORLQNDNTKFOADIASL-----NELEEA 828
890 EOLKEOLERNRDSPLQTVREKTLITELKLOOTLEEVKTLTQEKDDL-KOLQESLQIRRDQL 948
829 QNMLTEVQNSSEIVKLRQNHETAKIKELENTFEEMQREYDCLFNQLMESVQ-ENDAL 887
949 KSDIHTVNMNIDTOQLRNALSRLKQHTINTLKS--KISEVERNLMHEENTGETKD 1006
888 REBIKQRPSTSHVESMRSSGISDFFDEQKQDINLLHQFVQLSEV----- 932
1007 EFOKAVGIDKKODLEAKNTQTLTADVKNIEITEQOKRIFSLQEKNELOOMLESVIAEK 1066
933 -----QOIELQHSGISRLFRANQM-----KLDQSEPLKLCLESA----- 968
1067 EOLKTDLKENIEMTIENQEBELRLIGDELKQOEIVAOEKNAHAIKKEGELSRTCDRLAEVE 1126
969 EYIEEDNRQS-----DATEPICLKGFLKRRHQIKRLSQEHV--DMGEKRLDIIISQLE 1021
1127 EKLKESQOQLQOEQQOQLNVOESEMOMKINIEINELKNEKLTLEHMEVTERLELAQ 1186
1022 QEIEKSALMEATEATINEMREQNTNLESALLKSVIINKVYQORQIESLEKQNAEMTM 1081
1187 KLENVEEVKS-TYKERKVLKLOKSPETERDHLRG-----YIREIATGLQTKBELKIA 1240
1082 V-----YELQODRVTRSMSESRLR-VPPEDITLPGCPTSPSRQEQVATIKT----- 1129
1241 HIHLKHEQETIDELRRSVSEKTAQIINTOLEKSHTKLQEBEIPVILHBEQELLP--NVKKV 1298
1130 -----SITELOSQVSDINAELEN-----HLRQLKQDGNLART 1162
1299 SETQETVNNLELLTEQSTTKDSTTLARIEMERLRLNEKFOESQEBEIKSLTKEDNLTKI 1357
1163 QTDPEMSERCL-----SMEVRLAEDELDTKQKQLLDQRAQKSDDLCLI 1208
1358 KEALFVKHDQKEHIRETLAKIQSOSQOEGLANMKEDNETTKIVSEMEQFEP----- 1411
1209 -DQLOKNAQALVEQVHKATESLSLADAKPDQIILSSQYDSQIEKLNQLNNAKDELHDVR 1267
1412 --KDSALIRIEMIGLSKRLQESHDENKSVAKEDDLQRLQOEVLQSESQDLKENIKETIV 1469
1268 RIKODEISALRMEFL-----LQIETNEKENQAKFYAELQETKDYESNVABLK 1315
1470 AKHLEBEEEL-KVAHCCCLKQEBETINELRNLVNLSEKEIETI-----OKOLEAINDKLQNK 1524
1316 EKLQVETLSSVTVRCQAELEALKSAHKNISQAVEERNLIIVQHOAEMETIRETLKNG 1375
1525 IQEI---YEKEBQNLNIKOISVOENVNE-LKQFKEHRKAKDSALQIESKMLLETN----- 1576
1376 LAEASTQOSKQMEDAFRAINEVRATLMQNLQTKEDRDKGASKLEEVKTLQEQMINGRVR 1435
1577 -----RIQESQEBEIQIIMKE-----KEEMKRVQBALQIE---RDQLK-----EN 1612
1436 MSDTIAELEKTKAEQDLAVNKLTKDIBELKQCSKTQOLQWESLTRDQISFEIATIKK 1495
1613 TKEIVAKNK-----ESQEK-EYQFL-----KMTAVNETO-----EKMCETHEHKE----- 1651
1496 LELIVASSKGRRIIELEKCDQOQVLEBCKRLEKLSLESIKANKSEHSTCKMEKLOELQAE 1555
1652 -----QFETOKNLNE-NIETENIRLQIILHENLEEMRSVTKERDDLRSVETLKV---E 1701
1556 MKVLSNRNEKEKDPETKLETTFKTD-LLEVLEAKHQKVILYDVLVSQHERLKICLAE 1614

RESULT 6
Q9VK10 ID Q9VK10 PRELIMINARY; PRT; 2013 AA.
AC Q9VK10;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Cmet protein.
GN Cmet OR CG6392.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Berkely;
RC MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George K.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Randell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.D.G.,

Best Local Similarity	22.7%; Pred. No. 1.3e-32;	Matches	Conservative	513; Mismatches	1020;	Indels	598; Gaps	102;
7	VAVCVVRPLNSREESL-GETAQVYVWKTDNNVIVQVDSKSFNDRVFNHGNETTQNVYEE	65						
194	VOILRVPLNSWERSINGYRCLQOESSQCVAMIGPPEFQFDHVACETIDQETLFRV	253						
66	IAAPIIDSAIOGNYGTIFAYGQTASGKTYTWMGSEDHL-----GVIPRAIHDPQKI	117						
254	AGLPVENCLSGYNCSIFAYGQTGSGKTYTMLGVDGLEFKPSPNRMGMPRIFFELFARI	313						
118	KFPDR-----EFLRVSYMEIYNETITDLCGTQKMKPLIIRDDVNRNVYVADLDEV	171						
314	QAEESRRDERLYNCKCSFLBYINEQITDLL-----EPSSNQLQREDIKSGYVVENLICE	371						
172	VYTSEMALKWITGKESRRHYGTQKQNRSSRSHITFRMILES-R_EKGEPSNCEGSVKVSH	230						
372	VQSVQDILGIIQTGSLNRVVGATNMWRRSSHSVFTCVIESRWEKDSAN-----MRPAR	427						
231	LNLVDLAGSERAQAGAAVRUKRGCNINRSLFILGVIKKISQGVG--CFINVRDSSL	288						
428	LNLVDLAGSERQKTSAGEDRLKEAASINXSLTSTIGHVIMLVVDVANGPKHPIPRDSRL	487						
289	TRILQNSLGNPKTRIICTIP-VS-FDFTLTALQFASTAKMKNTPYVNVSVSTDEALLK	346						
488	TFLLQSLGNSKMTWIIANASPSVSCAAETLNTLKPQAKLIQNNVAVNEDSVL-----	545						
347	RYRKETMDIKKOLEEVSLETRAQMEKDL-----AQLEKOLLQKVQNEKIENL----	397						
546	ELRRQRLKKEELSLLKRONISFALSFGSATANFAESQVDSPPSVMHETGQQQAGNLIVY	605						
398	-----TRMLVTSSSLTLOQLKAKRK-----RRVTWCLGKINKMNSVADQFNPT	444						
606	ESGGCVMRMRKQLKSLEITLAGSLRHHVADASIKKLEABIEHLNRLVRQRED-----	659						
445	NITKTHKLSINLRE-----IDBSVCSGSDVPSNTLDTL-SEIEWNPATK	489						
660	-----TRSTKMLRFPREDKIQLBSLGNHISADSFLLENNVLSEETQLQAKIDKNPBLT	716						
490	LINQENISELSRADYDNLVDYQRLT-----EKEEMELKLE-----KNDLDF----	537						
717	RFALENIR-----LLD--QLRRFQFYEGEREILLGVSRLNRQLFQFLDE	761						
538	-----PALERTKK-----DQEMOLIHESINLKNLVKHREVTNQD	572						
762	NSDWQKHVDGIEFPQLKKTCTYELEKCRNLGSCLEENAKLSREINDLOQWVSDIRACTPD	821						
573	LENELSKVLLREKE-----DQIKKLOEYIDSQKL-----EN	605						
822	EHSVNVKQALLGTQNFEPHETLACEQANYVEIILQLDLVDOKIILDEERTLRGDTEA	881						
606	IKMDLSVLSIESID-----PKWQKT--LFDRAETVALDAKRSAFLRSENLELEKMKKE	657						
882	QAVRLKFDIEVLKDQLLLSKQQRNVYSELGETKSAVALESQNIILIQAEVRLRIKEN	941						
658	LATTYKQWENDIQLYO-----SOLEAK-KMQVDLEKELQSAFNEITKLTS	702						
942	YFELLKQQLDIPAMKSKQCDPKNDPAEUSEIDITFKMQASLEAKRLNMLYKSDIAS	1001						
703	LIDGKVPKDLLCN-----LELGKTIIDLOKELN-----KEVENEALREE	742						
1002	KACGDEEMDEVCKQAEATAEVI VCLQNEVLQKEVNDPQSKENVTEKQVEILETQMBE	1061						
743	VILLSLSLSEVERLAKELQDSKEELHIITSSKDKLFSEV-----VHKE	788						
1062	--LQDKLRTDMDNEQLOEQURGKMDMELLISINEMELLTSELBEILLNGHEGLTDACYQA	1119						
789	SRVQGLL-----EELGTKDKDLATTQSNYKSTQDFQNFKTLHMDPFQKYKMWLEEN	840						
1120	DLISGLDPDKRIWISQVGGILRTLSERELMIEDLESCLDANKKRCDIESMLSLKGAA	1179						
841	ERMNOEIVNLKSAEQFDSIGALKTELUSYTKO--ELOKTRVQREL-----	886						

QY 484 WNPATKLLNOENIESELNADYDNLVDYEQRLTEKEEMELKUKKXNDLDFEALREK 543
 Db 409 LRPSPKLTNLPKLPORRGIAKPKAGICKTLKEKLTQDNMD----- 450
 QY 544 TKQOQEMQIHIISNLKNIWKHREVNQDLEN-----ELSKV-----ELLREKEDQ 590
 Db 451 TPGRAKQLGRETASRIPSWMSKKYQESVPNCADPQTEISALTASNOVAKETIEKYBEQ 510
 QY 591 IKKLQBYIDSQKLENTKMDLSYLSLESIEDPKQMKQTLFDAETVALDAKESAFLESENLE 650
 Db 511 VRLKETIERLEMENKA-----VN 530
 QY 651 LKEKMKELATTKYQKMDNDIQLYQSOLEAKKQMOVDLEKELQSAFNBITKLTSLIDQKVK 710
 Db 531 LGEOPETHKAKSQMEEL---LSSISEKDSITVLSQSL-----ELSR 572
 QY 711 DLLCNLELGGKIDLOKELNKEVEE---NEALREVIL-----LSEIKSLPSEVERLKEIQ 764
 Db 573 DVLRSK-EDQMSMCPLESSCERICNKCLELERLLPLASAGLDSVACQFQDLRSEIA 631
 QY 765 DKSEELHIITSEKDLFSEVVHKEKSRVQGLLEIGIKTKDPLATTQSNYKSTDOEFQNFK- 823
 Db 632 ATMKLESMLSTSFHASCESVQKTTDCRLSEQISTAHDDFGQLQEKYNNLKHKWSQKL 691
 QY 824 ---TLHMF---PQYKMWLEENE---RMNQIVNLSKEAQPDSPSLGALKTELKSVKTQ 873
 Db 692 AIDTMQVDYNTIOQKYLQDEYRHLERLRSDEQOQLODENSKLQABEIGTLKE-----RVE 747
 QY 874 ELQKTRV---OERLNEMOLKEOLENRSPLQ-----TVREKTLITEKLOQLTLEVK 926
 Db 748 EIHSELLEVPNPDHEDMELQOELKRLSKLQWEPDEIOLNVECSNLMSTIQECDA 807
 QY 927 LTQE-----KDLKQLOESLIQIERDOLKSDIHDVTN---MNDTQOELRNALLESKQ----- 975
 Db 808 LREBKORTNDSLESKSS-----GVGTECDSPENELDTDLQOQTKLSKSQIQLT 861
 QY 976 -----HOETINTLKS-KISEVSRNLHMEENTGETKDEFOQKMGVIGDKKODLE 1022
 Db 862 DYSGRRLFYNAEQDQSPSLKLCLEPAKYL---EGDQKHQDASDSVFL---KGLK 914
 QY 1023 AKNTQTLTADVKNENIIEOQRK---IFSLIQKNELOOMLESVIAEKEQLTKDLKENLEM 1079
 Db 915 QRFQIVKINGEONLVKEEDMRDIIIFLQKQVDGKKNLIE-----EKEVINNNRAQITS 970
 QY 1080 -----TLENQO-ELRLAGDLKQOEIVA---QSKNHAIKKEGELSRTCDLAEVVEKIK 1130
 Db 971 LNOETIKQNAKTKICEELQYQDTQVOTANKQESQEVLTIKTSLAHLKSKVCELQKKLE 1030
 QY 1131 EKS---QOQLEKQOQLNVOEEMSEMOKKINIEMLK-----NELKNKELTLEHMET-- 1179
 Db 1031 KQSEDEKISELQSDIGETISECCLSMELKLADIVNWQAEELRPLDQLOESGVELQHHSTA 1090
 QY 1180 -ERLELAQUNENVEVKSITKERVILKELOKSPETERDHLR-----CYI 1223
 Db 1091 BESLNVEKPIQOETERTLTITYERRI-BQLESIQRAQOEELILEKRTDENKSLQLYM 1149
 QY 1224 REIATGLQTEELKIAHILHKEQETIDELRRSVSEKTAQIINTQDLEKSHKSLQEBIP 1283
 Db 1150 AKIETSENENSKFRAYCLDLKETQRYEBEOLQOTNEKLAVT-----TQOQVHLD 1200
 QY 1284 VLHBEQELLPNVKKVSETQETMNELELITEOSTTKDSTTLARIEMRLR---LNEKFSQS 1341
 Db 1201 VIKRSIQ-----EKITQAEKERNEL-----AVRHKAELKIRETLKEKESKY 1243
 QY 1342 BEIKSLTERDNLTKIKALEVVKHQLKEHIRETLAKIOESQOESQOSINMKEKDNETTK 1401
 Db 1244 EKLRQAEERD---KEISR-LEV-----MENTAELUKINSREVEL-----EGVK 1285
 QY 1402 IVSEMEQKPK---DSALIRIEIEMGLSKRLQSHDEKMSVAKK-DDLQRL----- 1450
 Db 1286 -----MEKQKLYDKSML---ELEQLOCTAD-QKSSDLLPGSSNENIDDLQKKQCVQD 1338
 QY 1451 QEVLOSQESQDKENTKIEIVAKH-----LETEELKVAHCLKEQBETIN 1494

Db 1339 LELLGEXKALLSELOKINGQHSNTIKKLEETEAEMITLTQKLELE---RCETAEKLETFK 1396
 QY 1495 ELRVNLSKETETISTIQOLEA-----INDKLQNKIOEIVYKEBQ 1534
 Db 1397 SKEADIKK---ALHCAQLRLHAYDKLVCEYERLKGCLSDNSKNLSENLOKKVERLHQAOLA 1453
 QY 1535 L-----NIKOI-SEVOENVNELKQPKHRKAKDSALQSTESKMLELTNRLOE--- 1580
 Db 1454 LOEGISGRDSEIKQURSELKDAIDENKTVREAKVGLNSLKAVOENWSAQEQFKQKIAD 1513
 QY 1581 ---SOBEIQIMIKEEMKRVQEQALQIERDQKLENTKEIVAKWESQOEYQFLKMTAVN 1637
 Db 1514 IKGSVDLQIKLSLQEV-----RDHLESNEELKRLKDAQELQNMVDMKCKLN 1563
 QY 1638 ETQKMKCEIHLKEQF---ETQKLNLENTETENIRLTQI-LHENLEMSVTKERDRLS 1693
 Db 1564 SS-----LREDPDKLEQTKLDLE---BOLRAKKEVIDRRSKELGEVTKDCENIRS 1610
 QY 1694 -----VEETLKVERDQK---ENLRITITRDLEKQEBELKIVHMLKHEQETIDK 1739
 Db 1611 DLEAQNDPFLKRETNLTITSLRLHNEQLLETSKNVLSDITAANNLNMKKNLHDLTK 1670
 QY 1740 ---LRGIVSEKTEINNMQKOLEHNDALKAQDLKIQELRIAHMLKQEQETIDKLRG 1795
 Db 1671 ECKSLRSDLSQSEYFQTKQLLDETISNLKEENRWEEKLSSGNKALKEDCE---KLRS 1727
 QY 1796 IVSEKTDKLSNMQKDLSENSNAKLOEK-----IQELKANEHQILITLK-----DVNE 1841
 Db 1728 TLESKELILOONKQOEELRTVINEKNGKNALLDAQKSNETAPTSURKAWIKQSLAIEA 1787
 QY 1842 TQKVSMEQO-LAKKOIKDQSLTSLKLEIENLAQELHENLEEMKSVYMKERDNLRR--- 1896
 Db 1788 ANKRSLEMEQKQVDRTRYEELSTLKTREINFRSEKERMDGTITSSLEDKRNLKELCT 1847
 QY 1897 VEETL-KLERDQKESLQETKARDLETQOELKFTARMLSKHEKTVDKLRKISEKTIQIS 1955
 Db 1848 VTELLAKLKEELPALHTQKVGSDVSI---ELNNGSNG-----SPTPAAPV 1889
 QY 1956 DIQKDLDP-KSKDELOKKIQELOKKEQLLQVKEVDNVMHKKINEMEQKKQF 2006
 Db 1890 ATKPLDCNSAECVPKSSSLTAETAE-----RNRMRMTAYDENRKOF 1930
 RESULT 10
 Q9NCF9 PRELIMINARY; PRT; 1931 AA.
 ID Q9NCF9
 AC Q9NCF9
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Mitotic kinesin-like motor protein CENP-ana.
 GN CANA OR CG4831.
 OS Drosophila melanogaster (Fruit fly).
 OC Rukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20351410; PubMed=10893249;
 RA Yucel J.K., Marszalek J.D., McIntosh J.R., Goldstein L.S.B.,
 RA Cleveland D.W., Philip A.V.,
 RT "CENP-meta, an Essential Kinetochores Kinesin Required for the
 RT Maintenance of Metaphase Chromosome Alignment in Drosophila.";
 RL J. Cell Biol. 150:1-12(2000).
 DR EMBL; AF20354; AAF32356.1; .
 DR HSSP; P56536; 2KIN.
 DR FlyBase; FBgn0040233; cana.
 DR GO; GO:000090; P:mitotic anaphase; IMP.
 DR InterPro; IPR001752; kinesin_motor.
 DR Pfam; PF00225; kinesin; 1.
 DR PRINTS; PRO0380; KINESINHEAVY.

1729 LESKELIIQONQLEELRLTVINEKNGKALLDAQLKSN--ETAPKSLQKAWIKOSLAIE 1786
 1928 TARMSEKHEKTVDK-----LRKISEKTTQISDIQKDLKSKDELQKKIQELQKK-- 1978
 1787 AANKSLEMEQWDRTRREYELRSLTKTREINFRSEKERMDDGTISSLLDKRNLEKLC 1846
 1979 ---ELQLLRVKEDVNMH--KKIN 1997
 1847 TVTEL-LAKIKRELPAHTQKN 1868

RESULT 11
 Q63714 PRELIMINARY; PRT; 3187 AA.
 AC Q63714;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
 DE GIANTIN (Golgi complex-associated protein of 364 kDa) (GCP364).
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 ON NCBI_TaxID=10116;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=KIDNEY;
 KC MEDLINE=98093490; PubMed=9431462;
 RA Toki C., Misumi Y., Fujiwara T., Sohma M., Nishiohara M., Ikehara Y.;
 RT "Identification and characterization of rat 364-kDa Golgi-associated
 RT protein recognized by autoantibodies from a patient with rheumatoid
 RT arthritis.";
 RL Cell Struct. Funct. 22:565-577(1997).
 CC -1- FUNCTION: MAY BE INVOLVED IN THE FORMATION AND/OR MAINTENANCE OF
 CC THE CHARACTERISTIC GOLGI STRUCTURE.
 CC -1- SUBUNIT: DISULFIDE-LINKED HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. GOLGI MEMBRANE.
 DR EMBL: D25543; BAA05026.1; --
 DR FIR; JC5837; JC5837.
 DR GO; GO:0005794; C:Golgi apparatus; IEA.
 DR GO; GO:0016021; C:Integral to membrane; IEA.
 DR GO; GO:0005871; C:kinesin complex; IEA.
 KW Golgi stack; Antigen; Coiled coil; Transmembrane.
 FT CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 1 3163
 FT TRANSMEM LUMENAL (POTENTIAL).
 FT DOMAIN 3164 3184
 FT DOMAIN 3185 3187
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 47 116
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 126 398
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 418 479
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 505 544
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 632 891
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 924 993
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 1010 1076
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 1103 1178
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 1321 1327
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 1348 1712
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 1765 2320
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 2331 2706
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 2356 2359
 FT POLY-GLU.
 FT DOMAIN 2723 2775
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 2800 2892
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 2953 3008
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 3069 3113
 FT COILED COIL (POTENTIAL).
 SQ SEQUENCE 3187 AA; 364295 MW; 48F035DF43647F51 CRC64;

Query Match 8.2%; Score 1089; DB 11; Length 3187;
 Best Local Similarity 20.9%; Pred. No. 9.1e-25;
 Matches 634; Conservative 558; Mismatches 1012; Indels 828; Gaps 124;
 113 IFQKIKKFPDRFLLRVSYMYINVTIT-----DLLCGTQKMKPLIIRVDNRVYA 165
 221 VMQRKLEHEEALLGRAQVVDLLQKELISAEQRNQLDLSQQLLEAEHSTLRNTMEA 277

166 DLTEBVVVTSEMAKWIITKGEK-----SRHYG-----ETKMNQRSS 201
 278 ERQESKILMEKVELEMAKKEELYLOQLERAGQAQAELEMOYGTLOQRHETEMEKK-- 335
 202 RSHTIFRMILRESKEGPNCEGSKVSHNLNVLGASERAAQTGAAGVRLKEGCNINRS 261
 336 ---TACISILQKNEQELQACD-ALKEENSKLLO-EQOEQAQAKAQAQLOQLEDE----- 384
 262 LFILQGVIKKLSQGVGGFINYRD-SKULTRILONSLGGNPKTRIITCTIPVFSFDELTAL 320
 385 ---LQOKSKEISQ-----FVNKPNLEKHETSSQTSL-----PDVYNEGVQAV 423
 321 OFASTAKYMNTPYVNEVSTDEALLKRVKRTMDLKKQLEEVSS-----LETRAQAMEKQD 375
 424 MEESVASLQKR--VLELENKCALLLSLELEELRAENEKLCRSRTLLLEAQNRAEAGD 480
 376 LA-----QALLEE--KDLLOKQVNEKIENITRLMVTSSSLT-----LQOEL 413
 481 MVCEVSTAGIALLNRSDSSTESGQDVLENTFSQHKELSVLLVEMKEAQEAEFAFKSQL 540
 414 KAKRKR-----RVTWCLGKINKMKNS-----NYADQF 440
 541 QOKRPGDYEVLDKRVQMMESGLPSVTARDVLC---APROKNSYPAVEGEQAGMRDQH 597
 441 NI-----PTNIT-----TKTHKL-----SINLLREI 461
 598 GTLEAGPLNDYGMELNSPQDGVDKSLSAHVCLCHQGELERLKTVLELETSLHTAKET 657
 462 DESVCSESDFVSNLTDLSEIENWPATKLNQ--ENIESELNLRADYDNLV-----LD- 513
 658 HEKNLSEK--AKEISSLAQL-----TKEFKESAEEARSTLTAVCEERDQLLYRVKELDV 709
 514 YEQRTKESEMELKLEKNDLDEFEALERKTKDQEMOLHIEISNLKLVKREVYNQDL 573
 710 LGELRAQVRELETSAB--AEKQGLDYESQRAQHNLLTEQIHSLSIEAKSKDVKIETL 766
 574 ENELSSKVELLEKEQIKKIQEYIDSKLENIMKDLSSLESIEDPKQMKQTLFDAETV 633
 767 QRELDGVQLOFSEGTQIKSLQSLQTKE-----SEVLEGAERMK----- 806
 634 ALDAKRESAFLRSENLELKEKMKELATTYQOMENDI-OLYQSOLEAKKKQVLDKELOS 692
 807 -----DISKEMBELSQAQSELEIAQMDQLLEKKKDVET-LQOITQE 849
 693 AFNEITKLTSLIDGKVPKDLLCNLELKGKITDQKELNKEV-----EENEALREEVILLSE 748
 850 KDOQV-----ELSFMTKRVQVNEEFKSLGVKLTKEQLSLSLR 891
 749 L-----KSLPSEVERLRKEITQDKSEELHIITSE 776
 892 ABGAKRQVEDSGAESPKHGPHESSAEEPVCKEALQOELEWLKESQKRLQAALIS 951
 777 KDLFSEVWHKESVQGLLEBIGTKDODLATTQSNYKSTDOEFONFKTLHMDFKQY--- 833
 952 RKELLQKVSLEELAKVREE--STKDSL--RESEKRELEEDSKN-----KDDPEKIGTS 1002
 834 -----KNVLEENRMQOEIVNLSKEAQKFOSSILGALKATSLSYKTOELQEKTRV-- 882
 1003 EWRELEVSRLTISEKEVELEGIRDLKEKAAAEELQALVQM---TQDLQNKTKIDL 1059
 883 -QERLNM-----POLKEQLNRRSPLQTV 906
 1060 LQEHITENQATIQKFIETMDAGDGSVAVKETSVPSPRAGGEGHWEPELEGK----IVDL 1116
 907 EREKTLITEKLOQTLLEVKTL---TOEKDDLKQLEQESIQIERDQKSIDHTVNNNDITQ 963
 1117 EKEKTLQKQLEALISRKALKKKAQEB--KHLKEELKEQKDAYR-HIQEFOFGQSKEN 1173
 964 EQLRNALSLKHOHQTINTLSKISEEV---SRNLHME-----ENTGETK 1005
 1174 ENIRAPLQOLQAKESTDQQLPQTGOQEPHSGEGLSLEGTPEASFDLHAAQSPHPGETA 1233
 1006 D-----EFQOKVMGID-KKQDLLEAKNTQTLTADV--KONEIIEQOKKIFSLQEKNE 1054

Db 1234 TQATVSVQIQDQLKIEVEKEELKISST-TSELTKSESVLLQEQINEQGLEIQN 1292
 QY 1055 LQOMLESVIAEKEQTLTKENIEMTIENQOEELRLIGDELKKQOEIVAQEKHAIKKEGE 1114
 Db 1293 LKAASHEAKAHTFOLQOEL-ESSQLKIALEHLKTLQPELTLQKHVGQKEEVSVLVQ 1351
 QY 1115 LSTCDRLAEVBEKKEKQSQ-----QLQEQOOLLNVQEMSMQKKINBIENL 1163
 Db 1352 LGEKEQTLTVQTEMEQERLIKALHTQLEMQAKEHEERLKQVQVEICLKQKPELE-- 1409
 QY 1164 KNELNKELTLEHMETERLELAQKLNNVEVKSITKERVUKHLQKSP-----E 1213
 Db 1410 -BESKAKOOLQKQALISREKALKENKSLQFOLSSARDAVEHLTKSLADVESQVSVQN 1468
 QY 1214 TERDHLRG-----YIREIBATGLQTK-----EELKIAHLKHEOHTIDELRR 1256
 Db 1469 QEKDALLGKALLQBERDKLIVEMDKSLLENOSLGGSCESLKLALGGUTDEKGL--MKE 1526
 QY 1257 SVSEKTAQIINTQDLEKSHTKLORIPVL-----HEEQELLPN 1294
 Db 1527 LESVRSKIAESTEWQEKHELOKEVEVLQSYENVSNBAERIQHVVESVRQEKQVYAK 1586
 QY 1295 V-----KKVSETQFTWNELE----- 1309
 Db 1587 LRSABSDKREKQLODAEQEMEMKMKRPAKSKQKKILEBENDRLRAEAQPVGA 1646
 QY 1310 -----LLEQSTTKDSTTLARIEMERLRLNEKFOESQOEIKSLTKERDNLTKTKEALEV 1363
 Db 1647 NESMEALLSSNASLKEE--LERITLEYTKLSEFALMAEKNTLSEETRNKLQVQAEQEL 1704
 QY 1364 KH-----DOLKEHTRE-TLAKIQESQSQEQSLNMKEDKNETTKIVSEMEQFKPKD 1413
 Db 1705 KOASLETTEKSDPKDVIEERTVAVVGSQEQDSLSSENAKLEDAEATILANSA---KP-- 1759
 QY 1414 SALLRIETEMGLSKRLQESHEDEKMSVAKEDDLO-RLQEV-LQESPD-----OLKENIKR 1467
 Db 1760 -----GVSETP-SSHDDINNYLOQLDQKGRITAELEMEKQKQDRELSQTLNEKN 1807
 QY 1468 IVAKHLETER-ELKVAHCLKEQBETINELRVNLSEKETEISTQKQLEALNDKLQNKIQ 1526
 Db 1808 ALLTQISAKDELK-----LEEEVAKINMLNQIQEELSRVTKUKETABEEKDDLEERL- 1862
 QY 1527 EYIEKHEQLNKTQISEVOENVNELQFKHEHRKAKDSALQSIESTKMLEITNRLQESQEBIQ 1586
 Db 1863 -----MNQLAELNGSIGNYQDVTDAIKN-----EQESEMQLKRCVSELEBEKQ 1909
 QY 1587 IMIKE-----KEBMKRVQEA-----LQIERDQKENTKEIVAKMKSQEKQYQFLK 1632
 Db 1910 QLVKEKTKVSEIKETWEIKIQGAKGPGSKIHAKELQELKKEQEVKQIQKDCIRYL 1969
 QY 1633 MTAVNETQERKCE-----IEHLKE---QFETOKLNLNLETENIR 1669
 Db 1970 RISALEKTVKALEPVHTESQKOLDATKGNLAQAVEHHKKAQAEUSSFKILLDDTQSEAR 2029
 QY 1670 LTQILHLENLEMSRVTKERDRLRSVEETLVKVERDQKLENRETITRDLEKQELKIVHMH 1729
 Db 2030 ---VLADNL-----KLKELQSNKESIK---SQIKQK-DEDLRLRLQEAEE-----KH 2070
 QY 1730 LKEHOETIDKLRGVSEKTEWEINMQKDLSEHSDALKQAQDLKTOBELRIAHMLKKEQET 1789
 Db 2071 RKEKKNQEKLDLHREKA-----HVEDTL-----AEIQVSLTRKDKDKMKELQQS 2115
 QY 1790 IDKLRGVSEKTKLNMQKDLN-----SNAK-----LQEKIOELKANEHLQILITLKKD 1838
 Db 2116 LDSTLAQLAFTKSMSSLODRDRVIDEAKWQRFQGDALQTKHEEVRLEKNECTALK-- 2173
 QY 1839 VNETQKVSMEQKQIQKQIDQSLTKSLKLEINLQAELEN--LEBEMKSVMKERNLNR 1896
 Db 2174 -DQLRQMTIHEELK-----ITVSKLEHD-----KEIWESKAQTELQHQKAYDKIQE 2220
 QY 1897 VEETKLKRLDQKLEQETKARDELQOELKTARMLSKHEKTVVDKLRE-----XI 1947

Db 2221 ENKELMSQLEAQAOLYHDSKNELTKLESELKSKLDQSTDKNLEKCRHENHNLGIIQ 2280
 QY 1948 SEKTIQ-----ISDIQKDLKSKD-----ELQKTIQELQKKEQLLRVKE-----DVNM 1991
 Db 2281 QEADIQNCXKCNCEQLETDLTASRELTTRLHDEINVKEQKIISLLSGKEAIVAIABELHQ 2340
 QY 1992 SH-KKINEMEQKKQPEPNYLCKCENDFQLTUKLHESLEERIVAKER-----DELR 2043
 Db 2341 QHSKETILENLSQEEENL--TLEENKRAVEKTNQLTAELETIKKESIQEQAQLDSFV 2399
 QY 2044 RIKESIKEMERDOFIATLRMIARDRQNHQVKEPKLLSD--GOQHLMESIREKCSRIKEL 2101
 Db 2400 KSMSSLODDRDRIVSDYROL-----BERHLSVILEKDELITQDAARENKLKBE 2447
 QY 2102 LKRYEMDDHYECLNRLSLDLKE-IEFHRIMKGL-----KYVLSVYTVKIKEROHE 2151
 Db 2448 IR--GLRGHMDIINSENAKLDAELIQVRRDLNEVITIKDSQORQLLELAQLOQNKELRNE 2504
 QY 2152 CINKPEMDFIDBEKQKELLIKIQLQOCDVPSRELRLDKLNQMDLHIEILKDFSES 2211
 Db 2505 CV-KLEGRLKGEABKQSLQMSLDALQENQGLSKEIKSFK-EQLTALHEEGALAVY-HA 2561
 QY 2212 EFPISIKTEBQOVL-----SNRKEMTQFLEEMLNTRFDIEKLKNGIQKENDRICQVNPFPN 2267
 Db 2562 QLRVREEVQKLTALSSSQKRTVDLQEBELVCVQKESKVKSEIEDKLR--ELKHLHN 2619
 QY 2268 RIILMNESTFEERSATISK---FWEQDLKSLKEKNEKLFNVTLTSLASGAQVNP 2324
 Db 2620 AGI-MRNETETAERVAELARDLVEMEOKLLTVTKENKDLTAQIAFGKSMSS----- 2671
 QY 2325 TODKNKPHVTSRAQTTEKIRELENSHEAKESAMHESKIIKMQKELEVTVNDIIAKLQ 2384
 Db 2672 LQDSRD-HATELSGL-----KKYDASLAKELAQKGRQDLGRESQVLSQA 2717
 QY 2385 AKVHESNKLEKTKETQVLQVLDKVALGAKPKVEIEDLKMVLKIDLEKMKNAKEFEKE- 2443
 Db 2718 FPL-----TTSANI-----SSRLKLNQOLISKDEQLHLSSESSH 2755
 QY 2444 ---ISATKATVEYQKQEVIRLLR--ENLRRSQQAQVTSVISEHTDPQSNKELTCGGSGI 2498
 Db 2756 NQVQSFKAQMTSLQNERDHLWNELEKFKSEBQSGSAA-----PSAASSPAE----- 2803
 QY 2499 VQNTKALI--LKSEHIREKESIKLQKQNEQLIKQKNEL-----LNNQHLSENEVTKW 2550
 Db 2804 VQSLKAMSSLODRDLKELKNLOQOYLQMNQOETELRPLKAQLOESQDTKALQMMK 2863
 QY 2551 ER-----TLKREAHKQVTCENGPKSPKVTGTASKKQITPSCCKERNLQDPV 2597
 Db 2864 BELRQENLSWQHEDQLRVEKNSWELHERRMKQOYLMAISDKQOQGLHQLNLREL---- 2919
 QY 2598 PKESPKSCFPDSRSKSLPSPHPVRYFNDNSSLG 2629
 Db 2920 -----RSSQTQILPTQYQORQASSG 2939

RESULT 12

Q8GVH3
 ID Q8GVH3 PRELIMINARY; PRT; 1967 AA.
 AC Q8GVH3;
 DT 01-MAR-2003 (T-EMBLrel. 23, Created)
 DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
 DE QJ1340-C08.3 protein.
 GN QJ1340-C08.3.
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 OX NCBI_Taxid=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA Sasaki T., Matsumoto T., Katayose Y.;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 09:30:02 ; Search time 47.178 Seconds
(without alignments)
2914.068 Million cell updates/sec

Title: US-10-045-631B-88
Perfect score: 13329
Sequence: 1 MAEGAVAVCVVRPLNSRES.....SQPGMHASSGKDVPCKTQ 2663

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	13313.5	99.9	2662	4	US-09-595-684B-31
2	3615.5	27.1	2954	4	US-09-150-867-1
3	1077.5	8.1	1388	4	US-09-572-191-2
4	1077.5	8.1	1388	4	US-09-723-262-2
5	1077.5	8.1	1388	4	US-09-723-219-2
6	1052.5	7.9	3878	4	US-09-914-259-11
7	1040.5	7.8	3248	1	US-08-353-700-1
8	1040.5	7.8	3248	5	PCT-US95-16216-1
9	1037	7.8	2482	1	US-08-328-254-6
10	946	7.1	1231	4	US-09-595-684B-23
11	941.5	7.1	1232	4	US-09-592-054-2
12	924.5	6.9	1234	4	US-09-592-054-8
13	889	6.7	1375	4	US-09-722-139-2
14	889	6.7	1375	4	US-09-721-832-2
15	889	6.7	1375	4	US-09-721-689-2
16	867.5	6.5	963	4	US-09-914-259-20
17	867	6.5	1898	1	US-08-056-200-94
18	867	6.5	1898	2	US-08-800-644-94
19	867	6.5	10182	4	US-09-134-001C-3159
20	866.5	6.5	957	4	US-09-914-259-16
21	863	6.5	956	4	US-09-914-259-17
22	856	6.4	967	4	US-09-914-259-21
23	848.5	6.4	963	4	US-09-914-259-22
24	845.5	6.3	1279	4	US-09-724-517-2
25	845.5	6.3	1279	4	US-09-641-807A-2
26	845.5	6.3	1279	4	US-09-723-096-2
27	844.5	6.3	975	4	US-09-914-259-19

28	844	6.3	1031	4	US-09-914-259-24
29	837	6.3	1027	4	US-09-914-259-27
30	835.5	6.3	1032	4	US-09-914-259-26
31	830	6.2	1066	3	US-09-541-782-8
32	830	6.2	1066	4	US-09-723-820-8
33	830	6.2	1066	4	US-10-270-085-8
34	824	6.2	2310	4	US-09-874-923-120
35	803.5	6.0	935	4	US-09-914-259-25
36	800.5	6.0	3696	4	US-09-134-001C-5080
37	782.5	5.9	1972	4	US-08-875-435B-3
38	779	5.8	1057	4	US-09-428-156B-2
39	778	5.8	1637	4	US-09-718-692-2
40	778	5.8	1637	4	US-09-718-852-2
41	778	5.8	1637	4	US-09-718-815-2
42	776.5	5.8	1056	4	US-09-595-684B-29
43	774.5	5.8	1972	4	US-08-875-435B-4
44	770	5.8	2101	1	US-08-466-390-4
45	770	5.8	2101	1	US-08-470-950-4

ALIGNMENTS

RESULT 1

US-09-595-684B-31
; Sequence 31, Application US/09595684B
; Patent No. 6544766
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Ohashi, Cara
; APPLICANT: Sakowicz, Roman
; APPLICANT: Vaisberg, Eugeni
; APPLICANT: Wood, Kenneth
; APPLICANT: Yu, Ming
; TITLE OF INVENTION: Human kinesins and methods of producing
; FILE OF INVENTION: and purifying human kinesins
; FILE REFERENCE: cytop036
; CURRENT APPLICATION NUMBER: US/09/595,684B
; CURRENT FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: 09/295,612
; PRIOR FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 2662
; TYPE: PRT
; ORGANISM: Human
US-09-595-684B-31

Query Match Best Local Similarity 99.9%; Score 13313.5; DB 4; Length 2662;
Matches 2662; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY	1	MAEGAVAVCVVRPLNSRESLSGETAQVYTKDNNVIYQVDSKSFNDRVPHGNETTK	60
DB	1	MAEGAVAVCVVRPLNSRESLSGETAQVYTKDNNVIYQVDSKSFNDRVPHGNETTK	60
QY	61	NYEETIAAPIIDSAIQGYNGTIFAYGQTASGKTYTMGSEDLHGVIPRAIHDFQIKKF	120
DB	61	NYEETIAAPIIDSAIQGYNGTIFAYGQTASGKTYTMGSEDLHGVIPRAIHDFQIKKF	120
QY	121	PREPFLRVSYMEIYNETITDLCTQKMKPLIREDVNRNVYVADLTTEVYVTSMAJK	180
DB	121	PREPFLRVSYMEIYNETITDLCTQKMKPLIREDVNRNVYVADLTTEVYVTSMAJK	180
QY	181	WTKGKSRHGVETKQNRSSRSHTIFRMILSRKGRPSNCEGSKVYSHLNLVDLAGSE	240
DB	181	WTKGKSRHGVETKQNRSSRSHTIFRMILSRKGRPSNCEGSKVYSHLNLVDLAGSE	240
QY	241	RAAQGTGAAGVRLKEGNCNINRSIFILGQVTKLSGQVGFINRYSKLTIRLQNSLGNP	300
DB	241	RAAQGTGAAGVRLKEGNCNINRSIFILGQVTKLSGQVGFINRYSKLTIRLQNSLGNP	300

QY	301	KTRIICITIPVSPDITLALQFASTAKYKMTNPVYNEVSTDEALLKRYRKEIMDLKKOLE	360	Db	1380	ESQSKQFSQSLNMKSKDNETTKIVSEMEQFPKDSALLRIEIMLGLSKRLQESHDEMKSV	1439
Db	301		360	QY	1441	AKEXDDIQRLOQEVLOQSESDQLKENIKIIVAKHLETEBELVAHCLKEQEBTINELRVNL	1500
QY	361	EVSLLETRAQAMEKDQALQLEEKDOLLQKVQNEKIENITRMLVTVSSSLTLOQELKAKKRR	420	Db	1440	AKEXDDIQRLOQEVLOQSESDQLKENIKIIVAKHLETEBELVAHCLKEQEBTINELRVNL	1499
Db	361		420	QY	1501	SEKETEISTIQKOLEANNDKLONKI OEIYKEGEOLNIQI SEVQENVNELQKQEHKRAK	1560
QY	421	VTWCLGKINQKNSNYADQFNIPNTNITTKTKHLSINLLREIDESVCSDESDFNSTDLTSL	480	Db	1500	SEKETEISTIQKOLEANNDKLONKI OEIYKEGEOLNIQI SEVQENVNELQKQEHKRAK	1559
Db	421		480	QY	1561	DSALQSTESKMLLETNRLQESOEI QIMIKEBEMKRVQEAQI QERDOLKENTKEIIVAKM	1620
QY	481	EIENWPATKLNQENIESELNSLRADYDNLVDYEQRLTEKEEMELKLEKNDLDEFAL	540	Db	1560	DSALQSTESKMLLETNRLQESOEI QIMIKEBEMKRVQEAQI QERDOLKENTKEIIVAKM	1619
Db	481		540	QY	1621	KESQEKYQFLKMTAVNETQKMCIEIHLKEQFQTKLNLNIENTENIRLTQILHENLEE	1680
QY	541	ERKTKKQDEMOLIHIEISNLKNLVGRVYNQDLENELSSKVLLREKDDOIKKQOYIDS	600	Db	1620	KESQEKYQFLKMTAVNETQKMCIEIHLKEQFQTKLNLNIENTENIRLTQILHENLEE	1679
Db	541		600	QY	1681	MRSVTKERDDILRSVEETLKVVERDQIKENLRSETITRDLKQEBELKIVHMLKHEQETIDKL	1740
QY	601	OKLENIKMOLSYSLIESIEDPKOMKOTLDAETVALDAKRESAFLSENLELKEKMKELAT	660	Db	1690	MRSVTKERDDILRSVEETLKVVERDQIKENLRSETITRDLKQEBELKIVHMLKHEQETIDKL	1739
Db	601		660	QY	1741	RGIVSEKTNELSNMOKDLEHNDALKAQDLKI QEBELRIAHMLKQEQETIDKLRGIVSEK	1800
QY	661	TYKOMENDIQLYQSOLEAKKQOVLDLEKELQSAFNEITKLTSLIDGKVPKDLLCNLELEG	720	Db	1740	RGIVSEKTNELSNMOKDLEHNDALKAQDLKI QEBELRIAHMLKQEQETIDKLRGIVSEK	1799
Db	661		720	QY	1801	TDKLSNMOKDLEHNDALKAQDLKI QEBELRIAHMLKQEQETIDKLRGIVSEK	1860
QY	721	KITDLOKELNEKEVERNEALREEVILLSELKSLPSEVERLRKBIQKQSEHLLHITSEKOKL	780	Db	1800	TDKLSNMOKDLEHNDALKAQDLKI QEBELRIAHMLKQEQETIDKLRGIVSEK	1859
Db	721		780	QY	1861	LTLKLEINENLAQELHENLEEMKSVMKERDNLRRVETTLKBRDOLKESIQETIKARDL	1920
QY	781	FSEVVKHSRVGCLLEIGTKTODLATTOSNKYSTDQEFQNFKTLHMDPEQKYKWLVEEN	840	Db	1860	LTLKLEINENLAQELHENLEEMKSVMKERDNLRRVETTLKBRDOLKESIQETIKARDL	1919
Db	781		840	QY	1921	EQQELKTAARMLSKSHKETVDKLRKIKSEKTIQISDIQKDLKSKDELQKKIQELQKKE	1980
QY	841	ERMNQEI VNLKSEAQFSSIGALKTLSYKTELQEKTRVQERLNEMEQKLEQENRD	900	Db	1920	EQQELKTAARMLSKSHKETVDKLRKIKSEKTIQISDIQKDLKSKDELQKKIQELQKKE	1979
Db	841		900	QY	1981	QLLRVKEDVNMGHKKINEMEQKQFEPNYLCKCEMDNPQLTKLHSHESLEEIRIVAKERD	2040
QY	901	SPLQTVREKTLITITEKLOTLBEVKTLTQKDDLKQLOESLQIERDOLKSDIHDVTNNMT	960	Db	1980	QLLRVKEDVNMGHKKINEMEQKQFEPNYLCKCEMDNPQLTKLHSHESLEEIRIVAKERD	2039
Db	901		960	QY	2041	ELRRIKESLKWEDQFIATLRMIARDRONHVKPEKRLSDGQOQHLMESLEKCSRIKE	2100
QY	961	DTQEQRLNALESIKHOETINTLTKSKI SEEVSNLHMEENTGTQDEFQKQVGVGDKQD	1020	Db	2040	ELRRIKESLKWEDQFIATLRMIARDRONHVKPEKRLSDGQOQHLMESLEKCSRIKE	2099
Db	961		1019	QY	2101	LLKRYSEMDDHYECLNRLSLDLEKETEPEHRIIMKKLYVLSYVTIKIEEQHECINKPEMDF	2160
QY	1021	LEAKNTQTLTADVKNONEIIEQQRKIFSLIQEKNELOQMLSVIAEKEQKLTDLKENIEMT	1080	Db	2100	LLKRYSEMDDHYECLNRLSLDLEKETEPEHRIIMKKLYVLSYVTIKIEEQHECINKPEMDF	2159
Db	1020		1079	QY	2161	IDEVEKQKELLIKIQHLOQDCDVPSPRELDRDLKLNQNMDLHIEILKDFSESEPPSIKTEF	2220
QY	1081	IENQEBELRLGDBELKKQOEIVAOEKONHAIKKEGELSRTCDRLAEVBEKLEKSOOLQEKQ	1140	Db	2160	IDEVEKQKELLIKIQHLOQDCDVPSPRELDRDLKLNQNMDLHIEILKDFSESEPPSIKTEF	2219
Db	1080		1139	QY	2221	QOVLNKRKEMTQPLEEWLNTFRDIEKLKNGIKQENDRI COVNNFFNNRIIAINNESTEFE	2280
QY	1141	QOVLNQOEESEMKKINEINENLKNELKNKELTLEHMETERLELAOKLNEYEVKSITK	1200	Db	2220	QOVLNKRKEMTQPLEEWLNTFRDIEKLKNGIKQENDRI COVNNFFNNRIIAINNESTEFE	2279
Db	1140		1199	QY	2281	ERSATTISKWEQDLKSLAKENKLFKNYQTLTKTSLASGAQVNPPTQDNKNPHVTSRATQL	2340
QY	1201	ERKVLKELQKSFETERDHLRGYIREIBATGLQTKGELKIAHILKHQETIDELRSVSE	1260	Db	2280	ERSATTISKWEQDLKSLAKENKLFKNYQTLTKTSLASGAQVNPPTQDNKNPHVTSRATQL	2339
Db	1200		1259	QY	2341	TTTEKIRELENSLHAEKESAMHESKIIKQKKELEVTNDIIAKLOAKVHESNKCLEKTKET	2400
QY	1261	KTAQIINTQDLEKSHTKLOBEIPVLEHEOHLNPNKVSQETQMTNLELLTTSQSTTKDS	1320	Db	2340	TTTEKIRELENSLHAEKESAMHESKIIKQKKELEVTNDIIAKLOAKVHESNKCLEKTKET	2399
Db	1260		1319	QY	2401	IOVLQDKVALGAPYKEEIEDLAKMLVKIDLEKMKNAKEPEKEISATKATVEYQEVIRL	2460
QY	1321	TTLARIEMERLRLNEKQESQEEIKSLTKERDNLTKTKEALEVVKHQDLKSHIRETLAKIQ	1380	Db	2400	IOVLQDKVALGAPYKEEIEDLAKMLVKIDLEKMKNAKEPEKEISATKATVEYQEVIRL	2459
Db	1320		1379	QY	2461	LRENLRSSQAQDTSVISEHTDPQSNKPIITCGGGSGIVONTKALIIKSHIRILEKISK	2520
QY	1381	ESQSKQFSQSLNMKSKDNETTKIVSEMEQFPKDSALLRIEIMLGLSKRLQESHDEMKSV	1440				

```

Db 2460 LRENLRSSQAQDTSVISEHTDPQSNKPLTCGGSGIVQNTKALILKSHIRLEIKS 2519
Qy 2521 LKQONEQLIKQKNELSNNOHLSNEVTKWERTLKREAHKQVTCENSPKPKVTGTASK 2580
Db 2520 LKQONEQLIKQKNELSNNOHLSNEVTKWERTLKREAHKQVTCENSPKPKVTGTASK 2579
Qy 2581 KQITPQCKERNLQDPVPKESPKSPFDSRSKSLPSPPHVPYRFDNSSLGLCPEVQNAAG 2640
Db 2580 KQITPQCKERNLQDPVPKESPKSPFDSRSKSLPSPPHVPYRFDNSSLGLCPEVQNAAG 2639
Qy 2641 SVDSQCPMWHASSGKQDVPECKTQ 2663
Db 2640 SVDSQCPMWHASSGKQDVPECKTQ 2662

RESULT 2
US-09-150-867-1
; Sequence 1, Application US/09150867
; Patent No. 6645748
; GENERAL INFORMATION:
; APPLICANT: Wood, Kenneth W.
; APPLICANT: Sakowicz, Roman
; APPLICANT: Goldstein, Lawrence S.B.
; APPLICANT: Cleveland, Don W.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Plus End-Directed Microtubule Motor Required for
; FILE REFERENCE: 18557C-000110US
; CURRENT APPLICATION NUMBER: US/09/150,867
; CURRENT FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: US 60/058,645
; EARLIER FILING DATE: 1997-09-11
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2954
; TYPE: PRT
; ORGANISM: Xenopus sp.
; FEATURE:
; OTHER INFORMATION: Xenopus centromere-associated protein-E (XCENP-E)
; OTHER INFORMATION: member of the kinesin superfamily of microtubule
; OTHER INFORMATION: motor proteins
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (1)..(472)
; OTHER INFORMATION: kinesin like motor domain
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (473)..(2752)
; OTHER INFORMATION: rod domain
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (2753)..(2954)
; OTHER INFORMATION: tail domain
US-09-150-867-1

Query Match 27.1%; Score 3615.5; DB 4; Length 2954;
Best Local Similarity 32.0%; Pred. No. 5 4e-162;
Matches 99%; Conservative 572; Mismatches 932; Indels 615; Gaps 86;

Qy 1 MABEGAVCVVRPLNSREESLGETAQQVYWKTDNNVYQVDSKSFNDFRVPHGNETTK 60
Db 1 MSEGDAVKCVVRPLIQREQ-GDQANLQWKAGNNTISQVDGTSKSFNDFRVNSHESHS 58
Qy 61 NYVEETAAPIIDSAIQYNGTTIPAYQGTASGKTYTMGSEDHGVITPRAIHDFQIKKF 120
Db 59 QIYQETAPVIRISALQGYNGTTIPAYQGTSSGKTYTMGTPNSIGIIPQAIQEVFKIQR 118
Qy 121 POREFLRYSMEIYNETITDLCGQTKMKPLIREDVNRNVYVADLTVEVYTTSEMALK 180
Db 119 FNRFLRYSMEIYNETIVKDLCCDRKKPLIREDVNRNVYVADLTVEVYVPHVIQ 178
Qy 181 WITGKESRHYGETKMNORSSRSHTIFRMILESREKGEPS---NCEGSKVKVSHLNLVDLA 237

```

```

Db 179 WIKKGEKNRYGETKMDHDSRSHTIFRMIVESDRNDPTNSENCDGAVVSHLNLVDLA 238
Qy 238 GSERAAQTGAAGVRLKEGCNINRSLFILGVIKKLSGQVGGFINYRDSKLTRILQNSLG 297
Db 239 GSERASQTGAAGVRLKEGCNINRSLFILGVIKKLSGQAGGFINYRDSKLTRILQNSLG 298
Qy 298 GNPKTRIICTITVPSDETLTALQFASTAKVMKNTPYNEVSTDEALLKRYRKEIMDKK 357
Db 299 GNAKTVIICTITVPSDETLTALQFASTAKVMKNTPYNEVSTDEALLKRYRKEIMDKK 358
Qy 358 QLE--EVSLETRAQAMEKQOLAQLLEKOLLQKVQNEKIENLTMLVTSLSLTQLQELKA 415
Db 359 QLENLESSETKAQAMAKEHTQLLAEIKOLHREDRIWHLTNIVVASSQES--QQQRV 417
Qy 416 KRKRVTWCLGKINKMKNINQDN---IPTNITTKHKLINLREIDSVCSBDV 471
Db 418 KRKRVTWAPKIQNSLHASGVDFDMLSLPGNFKAKFSDMPSPFPEIDDSVCTEFS 477
Qy 472 FSNLTDTLS---EIEWNPATKLNQE----- 494
Db 478 FDDALSMDNSGIDAENWNLASKVTHREKTSLHOSMIDFGQISDSVQPHDSSKENQLYLP 537
Qy 495 -----NIESE 499
Db 538 KUGSDMAECRKASFEKEITSLOOQLOSKEEKKELVQSPFKTAELEEQLSVAKNDEM 597
Qy 500 LNS----- 502
Db 598 TNSRHSINAEVQTDVEKEVVRKEMSVLGDGSGYNASNDLQDSSVQPKRLSSSHDECI 657
Qy 503 -----LRA 505
Db 658 RQMLEOKIVDLEEFENLNKKGENDKQSSQDFMESIQLCERALMAEKANALEALMRD 717
Qy 506 DYDNVLVDYEQLRTEKEEMELKKNLDLDEFEALERTKTKDOEMOLHEISNLKNLVKH 565
Db 718 NFDNIIENETLKRBIADLERSKENQETNEFEILEKTEKQKEHAQLIHEISGLKLVEN 777
Qy 566 REVYNODLENELSSKVELLREKEDQIKLQEVIDSQLENINMDLSVSL-----ESIEDPK 621
Db 778 AEMYNQNLSEDLTKTKLLKEQEIQLAELRKADNLQKVRNFDLSVSGDSEKLCERIF 837
Qy 622 QMKOTLFDATVADAKRESAFLRSENLELKEKMLATYTKOMENDIQLYSQLEAK-- 679
Db 838 QLKQSLSDAEAVTRDAQKCECSFLRSENLELKEKMDTNSWYNQKEKAAALFEKQLETS 897
Qy 680 --KMQVDLEKELQSAFNEITKLTSLIDGKVPKDLCLNLEKGITDLOKELNKEVENE 737
Db 898 NYKKMEADLQKELQSAFNEINVINGLAGKVPFDLLSRVELEKVSFQKLEKALEKN 957
Qy 738 ALREEVILLSELKSPSEVERLRKBIQDKSEELHIITSEKDLFSFVVHKESRVQGLLEE 797
Db 958 ALENEVTCLSYKFLPNEVECLKNQISKASEEIMLLKQSGEHSASIIISKEIIMQSQSEQ 1017
Qy 798 ICKTKDGLATTOSNYKSTDOEFQNKTLHMDPEQKVMVLENERMNOQIVNLKSAQKF 857
Db 1018 ILQLDEVHTHOSKVOQTEEQYLEMKKMHDDLFEXY-----IRNKSEADLLREMN 1070
Qy 858 DSSLGALKTSLSYKQELQEKTRVQERLNEMLKEQLEN--RDSPLQTVREKTLITE 915
Db 1071 KGTMSVEVKIADTKHELEETIRDEQLLHEKKYFFQAMQTFIPPTPLSDSIPPSKLVG 1130
Qy 916 KLQQTLE--EVKTLTQEKDDLQLOESLQIERDQLKSDIHTVNNNIDTQELRNALES 973
Db 1131 NSQDPIEDVHNLTALATERNNIMVCLTERNSLKEQV-----IDLNTQLQ-SLQQA 1182
Qy 974 KQQTINTLKSISE-EVSRNLHMEENTGETKDFQKQMGVIDKKQ--DLE-AKNTQTL 1029
Db 1183 STIEKSLQPKQDLKEGEVVKLLLEMLLKGHLTD---SLSIEKLQLENLVTETKQLT 1238
Qy 1030 TADVKNDEIIQQRKIFSLIOFKNELQOMLESVIAKEQLKTDLKENLMTLENQELRL 1089

```

1239 QEBMKNITI-----ERNELQTNFEDLKAEDSLKQDLSNIBOSIETQDELRA 1286
1090 LGDELKQOEIVAEQKHAIK-KEGEISRTCDRLAEVEEKLKESQOLQEQKQOQLLNQVE 1148
1287 AQEBELREKQOLVDSFRQOOLDCSVGISPNHDAVANQEKVSLGEVNSLOSEM-----LRG 1341
1149 EMSQMKK-----INEIENLKNELKNKELTLHEMETERLELAQKLNENYEVKSTIKERKV 1204
1342 ERDELQTSKRALVSELELLRAHVS-----VGENLEITKLNGLKEKEITLGKSESSEV 1394
1205 LKELOKSFETPRDHLRGV-----IRETEATGLQYKEELKTAHILKHEQETID 1252
1395 LKSMLENLKEDNNLKQAEYSSKENQFSLSEVFGSKLVDEIVLKAOLKAAERKLE 1454
1253 ELRSVSEKTAQIINTQDLE-KSHYKLOEIPVLHHEQELLPNVKVSETQETWNELELL 1311
1455 IKRDYFE-LVQTANTNLVEGKLETPLOAD-----HEED-----SIDRRSEEM-EIKVL 1501
1312 TEQSTTKDSTTLARIEMERLRINKEFPQSEBEIKSLTKERDNLKTIKALEVKGHDQLKEH 1371
1502 GEK-LERNOYLLERLQEBKELSNKLEITLOKEMETSVLLKODLOQLKESLSSENIILKEN 1560
1372 IRETL-----AKIOESQKOBOSLANKEKDNETTKIVSEMEQPKPKDSALLRTEIEML 1424
1561 IDTTLKHSDDTQAOLOKQOQLKQLOLAKNLAIAASONCPITQKE-----TSADCVHPLLEKIL 1617
1425 GLSKRLOESDBMSVAKEDDQLOLQVLOSQD-----KENIKETIVAKHLETEE 1477
1618 LLTEELHOKTNEQEKLLHEKNELEQAQVELKCEVEHLMKSMIESKSSLESIQEKHDTEQ 1677
1478 ELKVAHCCLEQEBETINELRNVLSKE-----TEISTIOKQLEBANDKONKIQEIYEKBE 1533
1678 QL-----LALKQMQVVVTQEKLELOQTHELTAEDVHLKENIE-LGLNFKNEAQQTKTKEQ 1732
1534 QL-NIKQISEQVENNV-ELKQFKEHRKAKDSALQSI---BSKMLELNRLOESDEEIQI 1587
1733 CLINENKLEQSQHLRQOEIEELMKSLKDKSALETLKESEQKVINL-----NQEMEM 1785
1588 MIKEEEMKRVQEAQOIERDQKENTKEIVAKWESQE---KEYQFLKMTAVNETQKMC 1645
1786 VMLEMEELKNSORTVIAERDQLODDLRSEVMSIETQDRLKAQEAQKQKQVQELTSQ 1845
1646 IEHLKEQFETQKLNLIETENIRUTQILHEWLEEMRSVTKERDRLRSVEETLKVVERQL 1705
1846 ISVLQEKISL-----LEN-----QMLY-NVATVKETLSRRDDLNSQKHLFSEIETL 1891
1706 KENLETTITRDLKQOEELKIVHHLKHEQETID---KLGRIVSEKTNESNMOKDLEHNS 1762
1892 SLSLKEK-EFALQEAQKDK-----ADAARKTIDITEKISNIEEQLLOQAINLKEITL-YER 1944
1763 DALKAQDLKIQEBELRIAHHLKEQOETIDKLGRIVSEKTDKLSNMOKOLENSAKLQEKI 1822
1945 ESL-----IQCKQLALNTEHLRETLKSLDALGKMEQERDEAANKVIALTERKWSLSLEQI 2000
1823 QE-----LKANEHOLITLKQVNETQKVSSEMOLKQIKQOGLITSLK-----EIE 1869
2001 NENVITLKEGEKETFYLRPSKQOSSQMBELRESLTKDLOLEAEKEISEATNEIK 2060
1870 NL-----NLAQELHENLEEMKSVMKERDNLRVVEETLKLERDQKESLQETKARDLEIQ 1924
2061 NLUTAKISSLEETLQNASILNEAVSERENLRUSKQOLVSELQSLUTL---KSRDHAFQ 2117
1925 ELKTAARMLSKHEKTVDLKREKISEKTIQISDIQDLDK---SKBELQKKIQELQ----- 1976
2118 -----SKREK---DEAVNKIASLAEBEIKULTKEMDEFDRDSKESLQCSHLSLELCT 2166
1977 -KKELOLLR-VKEDVNMMS-HKKINEMEQLKQFEPNYLKCENMDNFKLTKLHESLEIR 2033
2167 YKTELQMLKQOKEEDINNKAERKVEYDEL-----LQHLSLSSKEQLDQIQ 2210
2034 IVAKERDELRRIKESLQME-RDOFIATLEMTARDQRNHQVKEPKLLSDGOQHLMESLR 2092
2211 MELR-NEKLRNLYELCEKNDIMEKISVLRLM-----QNEPQOE---EDDVAERMDILE 2259

2093 EKCSRIKELLKRVSEM-DDHYECLNRLSLDLKEKIEBHR----- 2130
2260 SRNQEIQELMEKISAVTSEQHTLSSLSSELOKETEAKHCHMLNMIKESLSTLSRSGSL 2319
2131 -----IMKKLVVLSYVTKIKESQHCINKFEMDFIDEVEKQKELLIKIQLH 2177
2320 QTEHVKUNTOQLUNLKFVV--YRTAAVEDHSLIKDYEKDLAAEQKRDHDLQLOQL 2377
2178 QD-----CDVPSRELRLKLNQNMDL-----HTERILKOFSESE--FPSIKTFEQVLS 2225
2378 EQHGRKWSDSASELKECEIEFLNELLFKKANIISQVQDDSEFVQVFLNQVSGTLOEEL 2437
2226 NRKEMTOFLEWLNTRFDIEKLKNGIQKENDRIQOVANFFNFRNRIIATINVESTPEERSAT 2285
2438 HKGFQMWLEEFGLDHLVDKLSSEGMOQENNRRIASTTOLLTKRLKAVVQSKI---QREIT 2494
2286 I-SKEWQDLKSLKEKNEKLFKNYQTLKTSLSAQVN-----PTQDNKNPHVTSRA 2337
2495 VYLNQFEAKLOEKEQNKELMRMEHGPSVMEENARLLGILKTVQDE-----SKK 2548
2338 TOLTEKIRELENSLHBAKESAMKESKIKMOQKELEVTNDIIAKL---QAKVHESNCL 2394
2549 LQ---SRIKMLENELNLVKDDAMHKGKVAILODKL-LSRNAEAEALNAMQVLTCKKQDNL 2604
2395 EKTETIQVLODKVALGAKPYKEIEDLKMCLVKIDLEKMKNAKEFEKEISATKATVEYQ 2454
2605 QAAMKEITENLOKMAKAVPYKEIEDNLKTVKVIEMEKIKYSKATDOEIAVLKSLCDK 2664
2455 KEVIRLLRENRRSQQAODTSVISEHTDPQPSNKPFLTCGGSGGIVQNTKALILKSEHRL 2514
2665 EGBURLKEELRRQAQNDTIVCPKDYQKASTPVTTCGGSGGIVQSTAMVLQSEKAAL 2724
2515 EKEISKLKQONEQL-----IKQKNELLSNNOHLNEVKTWERTLKRAHKQV 2562
2725 ERELSHYKKYKHYHLRSTWSSSEDRKTKAKSDAHSHTGSHRSGSPHKTETVR---HGPV 2781
2563 TCE-----NSPK-----SPKVTGT-----ASK 2579
2782 TPERSEMPSLHLGSPKSESSTKRVVSPNRSEIYSLVMSPOKTGMHKLHILSPSKVGLHK 2841
2580 KKOITP-----SCKERNLODPVPKESPKSCFF 2607
2842 KRAISPARSEMPTQHVISPGKTGLHKNLTESTLFDNLSSPCQKQVQENL--NSPKGLF 2899
2608 DSRKSLSPHPVRVFDNSSLGLCEPVQAGAESVDSQGPWHASSGKDVPECKT 2662
2900 DVKSKSMFY-CPSQFQFNSKGLGDFSELNTAESNDKSAENMMWYEAKETAPECKT 2953

RESULT 3
US-09-572-191-2
; Sequence 2, Application US/09572191
; Patent No. 6355466
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Sakowicz, Roman
; APPLICANT: Wood, Kenneth
; TITLE OF INVENTION: No. 6355466el motor proteins and methods for
; TITLE OF INVENTION: their use
; FILE REFERENCE: 1017
; CURRENT APPLICATION NUMBER: US/09/572,191
; CURRENT FILING DATE: 2000-05-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1388
; TYPE: PRT
; ORGANISM: Human
US-09-572-191-2
Query Match 8.1%; Score 1077.5; DB 4; Length 1388;
Best Local Similarity 26.5%; Pred. No. 4.8e-43;

Query Match	8.1%;	Score 1077.5;	DB 4;	Length 1388;
Best Local Similarity	26.5%;	Pred. No. 4.8e-43;		
Matches	475;	Conservative 280;	Mismatches 521;	Indels 519; Gaps 79;
QY	3	EEGAVAVCVVRPLNSREESL--GETAQVYWKTDNNVYQVDS-----KSFNPD 50		
DB	23	EGDAIKVFIIRPPAERSGADGE-----QNCLSVLSTSLRLHNSPEPFTTFD 73		
QY	51	RVFHGNETNNVVEETAAPIDSAIOYNGTTIPAYQOTASGKTYTMWG-----SEDHL 103		
DB	74	HVADVDVTTQESFATVAKSIVSCMSGYNGTIPAYQOTGSGKTFMTMGPESDNFSHNLR 133		
QY	104	GVIPRAIHDF-----QKIKPPDRFLRVSMYIYNETITDLCGTQMKPLIIREDV 158		
DB	134	GVIPRSFEYFLSLIDREKAGAGKSFCLKCSFIEIYNEQIYDLL--DSASAGLYLREHI 191		
QY	159	NRNVYVADITEEVVYTFSEMAIKWITKGKSRHYGTYMQRSSRSHTIPRMILESEKGE 218		
DB	192	KKGVFVGVGAQEVVTSAAEAYQVLSGWRNRVASTSMNRSRSHAVFTTIESMEK-- 249		
QY	219	PSNCEGSVGVSHNLNVDLAGSRAAQTGAAGVRLKEGCNINRSLFLIGOVIKKLDGQVG 278		
DB	250	-SNEIVNIRTSNLNVDLAGSEKQKOTHAEGMKLKAAGNINRSLCLGQVITALVD--VG 306		
QY	279	G-----FINVRDSKLTRILQNSLGNPKTRIICTIPVS--FDTLTALOFASPTAKYMKNT 332		
DB	307	NGKQRHVCYEDSKLTFLLRDSLGNAKTAAIANVHFGSRFCGETLSTLFAQRAKLIKNN 366		
QY	333	PYVNE-----VSTDEALLKRYKEIMDLKK-QLEEVSLETR-----AQAEMKOLAQL-- 380		
DB	367	AVVNEDTQGNVSQLOAEVRLKELQALAEASGOTPPESFLTRDKKNTYMEYFOEAMLPFK 426		
QY	381	----BEKDLLOKVONEKIENLRMLVTSSSLTQQELKAKRKRVTCWCKINKMKNKN 436		
DB	427	KSQEKKSLIEKV-----TQLEDLTLLKKEKFI-----QSNKMIVKPR 463		
QY	437	ADQFNPTPTNTTKTKHLSI--NLRLREIDBSVCSDFSNLTDLTSE--IEWNPATKLLNQE 494		
DB	464	EDQI-----IRLEKLKHESRGGFLPEEQDRLSE--LRNEIQTLREQIEHHPVAKYAME 516		
QY	495	NISELSNLGRADYDNLVLDYQLRTEKEEMELKLEKNDLDEFEALERTKKQDEQQLTIH 554		
DB	517	N-----HSLR-----EENRRLLE-----PVKRAQEM----- 539		
QY	555	EISNLKNLVKRVYNOLENELSSKVLLREKEDQIKKLOEYIDSQKLENIKMDLSYSL 614		
DB	540	---DAQTIAKLEAF-----SEISG-----MEKSD--KNQGF----- 567		
QY	615	ESTEDPKMQQTTFDPAETVALDAKRESAFIRSENLELKEKMK--ELATTYKOMENDIQLYQ 673		
DB	568	---SPKAAQKEPCLFANT-----EKLKAQLLOIQITELNANSKQEYE 603		
QY	674	SOLEAKKMOVDLEKELQSAFNEITKLTSLIDGKVPKDLLCNLEKGTIDLOK--ELNKE 732		
DB	604	EFKELTRKQLESELOS-----LQCANLNL 631		
QY	733	--VEENALREEVILLSELKSLPSEVRLLAKEIQDKSEELHIITSEKDKLSEVVHVKESR 790		
DB	632	NLLEATKACKRQ-----EVSQLNK--IHAETLKIITP-----TKAYQLHSR 671		
QY	791	VOGLLEISGHTKODLATQTSNYKSTDOEFQNFKTLHMDFEQYKQWLEE--NERNMQEI 847		
DB	672	-----PVPKLSPEMGSGFLYT-----QNSSILDND-----ILNEPVPPMEQAF 712		
QY	848	VNLSEAKQFDSISGALKNTLSYKTQLOKTRVQERLNEEMEQLEQNLRSDPLOTVE 907		
DB	713	EAISEELRTVCEQMSALQAKL---DEEHEKNLKLQHVDKLHHSQWQELFS----- 762		
QY	908	REKTLITBKLOQTLEEVKTLTQEKDDIKLOLESLOIERDOLKSDIHD--TVNMNIDTQEQ 965		
DB	763	-----SFRIDWTQOQELLSQLNVLEKQLOET--QTKNDFLKSEVHDLRVVLHSAKDE-- 813		
QY	966	LRNALESILKQHETINTILKSISEVS--RNLMHEENTGETKDEQKQWVGIDKKODLEAK 1024		

```

RESULT 5
US-09-723-219-2
; Sequence 2, Application US/09723219
; Patent No. 6391613
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Sakowicz, Roman
; APPLICANT: Wood, Kenneth
; TITLE OF INVENTION: No. 6391613el motor proteins and methods for
; FILE REFERENCE: 1017
; CURRENT APPLICATION NUMBER: US/09/723,219
; CURRENT FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: US 09/572,191
; PRIOR FILING DATE: 2000-05-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1388
; TYPE: PRT
; ORGANISM: Human

```


Query Match 7.9%; Score 1052.5; DB 4; Length 3878;
 Best Local Similarity 19.5%; Pred. No. 2.2e-41;
 Matches 669; Conservative 603; Mismatches 1069; Indels 1089; Gaps 144;

Qy 26 TAQVYWKTDNNVIVQVDSKSFNDFRPHGN-----ETTKNVEYHIAPIIDSAIQVNGT 81
 Db 307 TMOISFLOEKIKVYMEQDK-----KVENSKBEIQEKETIEELNTKIIE----- 353
 Qy 82 IFAYGQTASGTYTWMG-----SEHILGVIPIRAIHDFQIKIKKFFDPRELLAVSYMEIYN- 136
 Db 354 -----EKKTLDELKDLTADKLLGLBLOEQIVQKNQOEIKRMK-----LELTNS 395
 Qy 137 -----ETITDILCGTKKPLIREDVNRNVVADLTBEVVYTSBALKWITFKGKS 188
 Db 396 KOKERQSSEELIKLMGTVEELQ-----KRNHSDSFETDIIVQMEQETQKLEQL-RAELD 450
 Qy 189 RHYGE--TKMQRSSRSHITFRMILESREKGEPSNC-----EGSVKVSHUNLVLD 236
 Db 451 EMYGQIVQMKQELIRQHMAQWEEEMKTRHKGEMENALRSYNSITVNEQIKLMVAINEL 510
 Qy 237 -----AGSERAAQTGAAGVRLKEGCNINRSILFILGVVKKLSDG-----QVGGFINVRD 285
 Db 511 NIKLODTSQKEKBEELGLILEKCALORQ-----LEDVLELTSFREIQIARQIARQIAEQE 567
 Qy 286 SKLITRILQNSLGGNPKTRIICTITP-----VSFDETLTALQFAPASTAKYMKNTPPYNEVST 340
 Db 568 SKL-----NEAHKSLSTVEDLKAIVSASERKELE-----LKHEAEVTNYKI 610
 Qy 341 DBALKRYRKEIMDLKKQLEVSLE---TRAQMEKQDQAOLLEKSKOLLQKVONKEIENL 397
 Db 611 KLEMLEKKNVLDRAESQEAELERLTQLLFSHEEELSKLEDLEIHRINIEK----- 666
 Qy 398 TRMLVTSSTLTQOELKAKKRVVWCLGKINKMNSVADQFNPTNITTTKHK--LSI 455
 Db 667 -----LKDNLGIHYKQOIID---GLQEMSQKIETMQFE-KDNLIITQONQILLEI 711
 Qy 456 NLLREIDES-VCSESDFVSNTLDTL--SETEWNPATKILNQENIESELNSLRADYNLVD 513
 Db 712 SKLQLOQLVNSKSEEMTLQINELQKEIE-----ILAQE-EKEKGTEQEVQEL-- 760
 Qy 514 YBQARTEKEEMBLKKEKNDLOB----- 536
 Db 761 --QLATLELKQWKEKE-NDLOKFAQLEAENSILKDEKKTLEDMLKHTPVSQERLIP 817
 Qy 537 FEALERKTK-----KQOEMQLIHEISNLKLV-----KHREYVNDLE 574
 Db 818 LDSIKSKSDSVWEKEIILIEBENEDLKQOQIOLNEEIEKQRTTFSAEKNFEVNYQELQ 877
 Qy 575 NELSSKVELLREKEDQIKKLOEYIDSKLENIMKMDLSIESIEDPK-QMKQTLFD----- 629
 Db 878 BEYACLLKVKDLEDSDS-KNKQELLEYKSKLALNEEL--HLQINPTTVKMKSSVDFEDKT 934
 Qy 630 --AETVAL-----DAKRESAFURSENLEKKEKELATYKQWENDIOLYQSOLEA 678
 Db 935 FVAETLEMGEVVEKDTTELMEKLEVTREKLELSQRLSDLSLEQKQKHGEISFLNEBVS 994
 Qy 679 KK-KMQVDLE-KELQSAFNE-----ITKLTSLDGKVPKLLCNLEEGKITDLQ 726
 Db 995 LKQKEQVSLRCRELEIINHNAENAVOSCDTQVSSLLDGV--TWTSGAGSVSKVN 1051
 Qy 727 KELNKE-----VEE-----NEALREBIVILLSKSL-----PSEVERLRKEI 763
 Db 1052 KSFGESKIMVEDKVSFENMTVGBESKQELILDHLPVSVTKESSLRATQPSNDKLQK-- 1109
 Qy 764 QDKSEELHITTSKDKLPSFVHVHESVQGLLEIEIKTKDDLATQSNYKSTQOEQNFK 823
 Db 1110 -----ELNVLSKQNDL-----RLQMEAORI-----CULSVYST----- 1138
 Qy 824 TLHMDPEOKYKWLLEENRMNQEIVNLSKEAQFDDSLGALKTSLSYKTOELQKTRVQ 883
 Db 1139 --HVDQVREYM-----ENEK-----DKALCSLKEELIFAQ---BEKIKELQ 1174

Qy 884 ERLNEMQLKQLENRDSPLQTVREKT-----LITEKLOOTLEE-----VKTLTQE 930
 Db 1175 -KIHOLE-----LQTMKTQETGDEKGLHLHLIGLKQKAVSECSYFLOTLCV 1221
 Qy 931 KDDLKQLQESLIQIERDQKSDIHDTVNMNIDTQBLRNALBSLKHQOFTINTLKSISEE 990
 Db 1222 LGYYTTPALKCEVNAED-KENSGDYISENED--PELODYRYEVQDFQENMHTLLNKVTEE 1278
 Qy 991 VSRNLHM-----BENTGETKDFQKQVGVGLDKKQDLEAKNTQILT 1030
 Db 1279 YNKLLVLQTRLSKIWGQOTQGMKLEFGEENLPKEBETFLSIHSQMTNLEIDVNHKSLS 1338
 Qy 1031 A-DVKNWEIIEOQKPIFSLI---QEKNELOQMLES----- 1061
 Db 1339 SLOLEKTKLEQVQOELESLISSLOQQQLKETEONVEAIEHCLQKRLQAVSESTVPPSLPV 1398
 Qy 1062 --VIAEKEQLKT---DLKENIEMTIE-----NOBEL----- 1087
 Db 1399 DSVVITESDAORTMYPGSCVKKNIDGTIEFSGFVGKBEETNIVKLEKQYQOLBEEVAK 1458
 Qy 1088 -----RLLGDE---LKKQBIIVAOEKNAHAKKEGELSRTCURLA--- 1123
 Db 1459 VIVSMSIAPAOOTELSRISGGKENTASSKQAHAVQOQBHYF---NEMKLSQDQIGFQTF 1515
 Qy 1124 -EVEEKLKESQOQO---EKOQQLNVQEESEMSEMKKINIE 1161
 Db 1516 ETVDYVFKBEFKPUSKELGEHGKEILLNSDPHDIPESKDCVLTISEEMFKDKTFFVRQ 1575
 Qy 1162 NLKNEKL-----NKELTLEHMETERLELAQKINENYEEVKSITKERVKLKLOKSFETE 1215
 Db 1576 SIHDEISVSSMDASQMLNHEQLE--DMRQELVRQYQEHQOQATQSSIDNE--NLVSE 1630
 Qy 1216 RDHLGTYTRETATGLQYKBEELKIAHILKHEQHTIDELRRSV-----SEKTAQIINTQ 1269
 Db 1631 RE-RVLLBELEA-----LKQLSLAGREKLCCELRNSSTQNGENQOGEVBOT 1678
 Qy 1270 DLEKSHTKLOEIP-VLHEEQE-----LLPNVKKVSETOETMNE--LELLTE 1313
 Db 1679 FKEKELDRKPEDVPPEILSNERYALQKANNRLKILLEVVKTTAAVEETIGHRVGLDOR 1738
 Qy 1314 QSTTKDSTTL-----ARI-----EMERLRJN-----EKF 1337
 Db 1739 SSKSQSSASLIWRSEASVKSVCVHEHTRVTDSPISYSGSDMPRNDNMMSKVTEGT 1798
 Qy 1338 QSQSBEIKS-----LTKERDNLK-TIKEALEVHKDOLKEHIREITLAKIQESQSKO--- 1386
 Db 1799 ELSQRLVRSFGAGTIDPENEEMLMLNISRLQAQAVEKLELEAISETSQLEHAKVTOTELM 1858
 Qy 1387 EQSLANKEKDNETTKIVSEM----- 1410
 Db 1859 RESFRQKQATBSLKQOEELRERLHEESRAREQLAVELSKASGVIDGYADEKTLFERQIQ 1918
 Qy 1411 PKDSALLRIEIMGLSKLOESHDMSKVAKEKDDLOBLQLOEVLOSES-----DOLKENIK 1466
 Db 1919 EKTDIIDRLQELLCSNLQLEAEAPQOQIQEERELLSRQKEAMKAEAGPVEQOLLQETE 1978
 Qy 1467 EIVAKHLET-----EELVAKCLCKQOEBETINELRVNLSEKETEISTIOKOLE 1515
 Db 1979 KLMKEKLEVQCOAKVRDLDQKQKALEIDVEQVSRFIELE---QEKNTLMDLQOQ 2035
 Qy 1516 AINDKLQNK-----IOEIVEKEEQLMI---KOISEVOENNVNKLQF 1553
 Db 2036 ALEKQLEKRRKFFIDQAIIDREHERDVQOEIQKLEQQLKVVVPFQFISEHQ--TREVEQL 2093
 Qy 1554 KEHRKADSALQSIKSMLELTLNRLQESOEI-----QIMKEKEEMKRVQOE-- 1600
 Db 2094 ANHLKEKTDKCSLLSKQLOQORDIQERNEEIEKLEFRVRELEQALIVSADTFQKVEDRX 2153
 Qy 1601 -----ALOIERDOLKENTYKEIVAKMESQOEYQFLKMTAVNETQEKWCE 1645
 Db 2154 HFGAVEAKPESLSEVQLOAERDAIDRKEKIT---NLEQLQEQFRE-----ELEKNKEE 2204
 Qy 1646 IEHLKEQFQTK-----INLENITENIRITQILHENLEEMRSVTKERDRLRSVEETL--- 1698

Db	2205	VOQLHMQLTQKKESTTRLOEBOEN	2266
Qy	1699	-----KVERDQLKENL	1730
Db	2261	KFAQIIQEKEVEIDQLANEQVTKLQQQLKIITDNKVIKEKNELRDLETOELCLMSDOECV	2320
Qy	1731	-KEHQEYIDKLRGVSEKTNESNM-QK	1775
Db	2321	KRNREEIEQLENEVIEKLQOELANIGOKTSMNAHSLSEEDADSLKHQDVLVIAEKLALEQ	2380
Qy	1776	LRIAH	1810
Db	2381	VETANEEMTFMKNVLKETNFMKNQLOTOELPSLKRERESVEKIOSIPENSVMVAIDHLKSD	2440
Qy	1811	LENSNAKLOKIOELKANEHQ	1853
Db	2441	KPELEVVLTE--DALKSLENOTPKPSFEENGKGSINLETFRLLOQUESTVSAKOLELTQCY	2498
Qy	1854	KQIKDQS	1889
Db	2499	KQIKDMQEOGQFETMLQKKIVNLOKIVKEKVAALVSQIQLAEAVAYAFCDQNOTISS	2558
Qy	1890	--ERDNLRRVEETLKLERDOLKESLOETKARDLEIQQ--ELKTARMLSEKHEKTVDKLR	1944
Db	2559	EPERTNIONLQ--LREDSLGDSISALTURISELESQVVMHTSLTLEKEQVEIAEK--	2613
Qy	1945	EKISETIQISDIQKOLDKSKDELQKIOBLQKELQKQLLVRKVEDVMNHHK--INEMEQ	2001
Db	2614	NVLEKEKKLLELQKLEGNKKQREKEKRSQDVEVLTKTTTVELFHSNEESGFFNELEA	2672
Qy	2002	LKKQFEPNYCKEMONFQUTKKLHLSLEIRIVAKE	2048
Db	2673	LRAE--SVATKAELASY--KEKAELQOE--ELLVKETNMNTSLQKDLQSVRDHLAEAK	2725
Qy	2049	LX-MERDOFATLREMIARDRONHOVKPEKRLSDGQOHLMESIRKCSRIKELLRYSE	2107
Db	2726	LSILEKEDETEVOESKACMFPELPKLSKSIAOSDTGTLKISSNQTPQI--LVKNAG	2782
Qy	2108	MDDHYECLNRLSDLEKIEBFHRIMKKLVVLVSVTKIEEQHECNKEMDFIDEVEK-	2166
Db	2783	-----IQINLQSECSSEVTELIISOFTBKIEKMQELHAEILDMESRHSISETEL	2832
Qy	2167	QKELLKIKIHLQODCDVPSRELRLKL	2197
Db	2833	KREHYVAVQLLKESCGTLKAVIQCLARSKEVFGFYNNCFSTLCDSGDWGGIYLTHSQGF	2892
Qy	2198	DLHIEILKDFSES--EFP	2237
Db	2893	DI--ASERGEESASATDSFPKTKIGLLRAVHNEGMOVLSTESPYSDGEDHSGIQQVSEFP	2951
Qy	2238	LNTR----FDIEKLNKGIOK-ENDRICQVNN	2270
Db	2952	LEERKAVINTISLKDILITKQOLORAEVYDSQSHEFSDFDWGCELLLALQOVLEERSV	3011
Qy	2271	AINNESTE	2314
Db	3012	LIAAFRELTALGTDDAVGLLNCLEQRIQEOGVEYQAAAMECLQKADRRSLSEIQALHAQ	3071
Qy	2315	LASGAQVNPPTQDNKP	2344
Db	3072	MNGRKITLKRQOSEKPSQELLEYNTQQKQSOMLENQVELSSMKDRATELQELSSERKM	3131
Qy	2345	IRELENSLHEAK-----ESAMHESKIKIMQK-----ELEVNTDIIAKILOAKVHES	2390
Db	3132	VABLKSSELAQTKLELBTTLKQAQHHKHEIAEFLEVKDXTDVEHLLNDTLTASEQKSR	3191
Qy	2391	NKCLEKTGTIQLQDKVALGAKPYKEEIEDLKMVLKIDLEKMQNAKEFEKEISATKAT	2450
Db	3192	QWALEKEKA-----KLGRSEERDKEELEDLKFSL--ESQKQRLQ-----LNL	3233
Qy	2451	VEYOKEVIRLLRENLRARSQAQDTSVISEHTDQPSNKPFLTGGGGSGIYONTKALILKSE	2510

Db	3234	LEQKQKLLNESQQKIESQRMLYDAQLSEE-----QGRNLEL-----QVLLSEE	3276
Qy	2511	HIRLEKEISKLKQONE-----OLIKQKNELLSNNQHLSNE	2545
Db	3277	KVTRREMSYTLDRERELHAQLQSSDGTGQSRPPLPSDDLKELQKLEEKHSRIVELLE	3336
Qy	2546	VKTWKERTLK-----REAHKQVTCENGSPKSPKVTGTGTAKKKQITPTSOCKERNLQDPV	2597
Db	3337	TEKYKLSLQTRQOMEXDRQVHRKTL-----QTEQEANTEGQKK-MHELQSKVEDLQRL	3390
Qy	2598	PKESPKSCFPDPSRSKSLPSHPVRYFNNSLGLCPVQNAQESVD-----SQ	2645
Db	3391	BEKQYQVYKLDLEGRLQ-----GIMQEFQKQELEREKREKRRILYQNLNE	3437
Qy	2646	PGPWHAASSGK	2655
Db	3438	PTTWLSLTSR	3447
RESULT 7			
US-08-353-700-1			
; Sequence 1, Application US/08353700			
; Patent No. 5599919			
; GENERAL INFORMATION:			
; APPLICANT: YEN, TIMOTHY J.			
; APPLICANT: RATTNER, JEROME B.			
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A			
; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,			
; TITLE OF INVENTION: AND METHODS OF USE			
; NUMBER OF SEQUENCES: 4			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: DANN, DOREMAN, HERRELL AND SKILLMAN			
; STREET: 1601 MARKET STREET, SUITE 720			
; CITY: PHILADELPHIA			
; STATE: PA			
; COUNTRY: USA			
; ZIP: 19103-2307			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: Floppy disk			
; COMPUTER: IBM PC compatible			
; OPERATING SYSTEM: PC-DOS/MS-DOS			
; SOFTWARE: PatentIn Release #1.0, Version #1.25			
; CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: US/08/353,700			
; FILING DATE: 09-DEC-1994			
; CLASSIFICATION: 435			
; ATTORNEY/AGENT INFORMATION:			
; NAME: REED, JANET E.			
; REGISTRATION NUMBER: 36,252			
; TELECOMMUNICATION INFORMATION:			
; TELEPHONE: (215) 563-4100			
; TELEFAX: (215) 563-4044			
; INFORMATION FOR SEQ ID NO: 1:			
; SEQUENCE CHARACTERISTICS:			
; LENGTH: 3248 amino acids			
; TYPE: amino acid			
; STRANDEDNESS: single			
; TOPOLOGY: linear			
; MOLECULE TYPE: protein			
; HYPOTHETICAL: NO			
; ANTI-SENSE: NO			
; ORIGINAL SOURCE:			
; ORGANISM: HUMAN			
US-08-353-700-1			

Query Match 7.8%; Score 1040.5; DB 1; Length 3248;
Best Local Similarity 19.6%; Pred. No. 6.7e-41;
Matches 635; Conservative 583; Mismatches 1033; Indels 985; Gaps 132;

Qy 154 IREDVNRNVYADLTTEVVYTSEMALKWTTTKEGRHYGETK----MNQRSGRSHITFRM 209
||| : : : : :
Db 237 IRDFGSAYFSFLE---VTPGRTSIQIGKRDNASFFGNSSPHLLDQLKAQNOLRNK 293
||| : : : : :

QY 210 I--LESREKGEPSNCEGVS-KVSHNL-----VDLAGSER-----AAOTGAA 248
Db 294 INELELRLOGHEKEMQVNFQBLQLEBAKAKVELIEKERVNLKCRDELVRTTAQYDQA 353
QY 249 GVRUKBGCNINRSFILGOVTKLSDGVGVFINVRSKLTRIILONSIGGNPKTRIICTI 308
Db 354 STKYT-----ALEQKLKLTEDLSCORQNAESARCS--LEQKIKERK 394
QY 309 TPVSFDETLTALQ-----PASTAKYMNTPVNEVSUDE--ALLKRYRK 350
Db 395 ---BFOBELSQOORSFOTLDOBCIQMKARLTQLOQAQKMNVLQAEIDLKLTYSVQOULEN 451
QY 351 BIMDLKKOLEVSLTRQAAMEKQOLAQLE-----EKDLLOKVQNEK-----TEN 396
Db 452 NLERFKQKLCRAEQAFQASQIKENELRSMEEKNNLLKSHSQKAREVCHLEAELKN 511
QY 397 LTRMLVTSSTLTQOBLKAKRRRVTWCLG---KINMKMNSYADOFNIPNITTKTHKL 453
Db 512 IKQCL--NOSQNFABEMKAKNTSOETMLRDLQEKINOQOENSITLEKKLAVADLEKQDRC 569
QY 454 SINLLREIDESVCSGESDVFSNT---LDTLGEIE-----WNPAT-KL 490
Db 570 SQDLKKGHEHHIQNDKLSKTEKESKALLSALELKKEVEELKEEKTLPSCWKSENEKL 629
QY 491 LNQ-----ENTIESELNSURA-----DYDNLVLDYQBLRTEKEEMELKLEKNDLD 535
Db 630 LTQMESEKENLQSKINHLETCLKTQOIKSHEYNRV---RTLEMDRENLSVEIRNLHNL 686
QY 536 EPEALERTKK--DOEMQLIHHISMLKNLVKREYVNOQDLENELSSKVELLREKEDQIK 593
Db 687 DKSVEVETOKLAYMELQAKAFSDQK-----HKEIENMCLKTSQLTQGVEDLEHK 738
QY 594 LQ---EYIDSKLBNIKMDLSYSESTEDPKOMKOTLFDATV-----ALDAK- 638
Db 739 LOLLNEIMDK---DRCQDLHAYESLRDLKSK-----DASLVNEDHQSRLAFDQOP 791
QY 639 -----RESAFLRSLENLEKMKELATTYKQMENTIOLYOSQLEPAKKQMVULEK 688
Db 792 AMHGSFANIGQGSMPSESRSECRLEADQSPKNSAILQNRVDSLEFSLESQKMNDSLQK 851
QY 689 ELQSAFNEITKLTSLDGKVPKDL-----CNLEEGKITDLOKEL-----NKEVERN 736
Db 852 QCE-----BLVQ-----IKGEIENLMKAQOMHQSFAVETSQRIKSQOEDTSAHQNVVAETL 903
QY 737 BAL-----BREVILLSELKSLPEVERLRKE---IQCKSELMIITS-----EKDLKFSEV 785
Db 904 SALENKEKELQLNDKVDVEQAEIOELKXSNHLLDSLEKQLQLSETLSLEKKEMSSIIS 963
QY 786 HKESRVQGLLEIGTKD---DLATQSNYKSTDOEFQNFKTLHMDFOQKYQWVLENER 842
Db 964 LNKREIEELTQENGTLKEINASINQEKMNLIQKSESFANY-----IDEREKSISELSQ 1017
QY 843 MNOEIVNLSKEAQKFDSSIGALKTELSTYKTOELOKTRV-----QERLNEMEQ 891
Db 1018 YKOEKLILLQRCBE---TGNAYEDLSQYKAAQEKNSKLECLNNECTSLCENRKNLEQ 1073
QY 892 LKEQLENRDSPLQT-----VEREKTLI--TEKLQOTLEEVKTLTQ-----EKDDLKQLOE 939
Db 1074 LKEAFAKEHOFELTKLAFABERNQNLMLELTVQALRSEMIDNQNNSSEAGGLKQETM 1133
QY 940 SIQIERDQKSDIHDTPVNNIDTQEOLRNALSLEKQHOETI--NTLKSQISE-----989
Db 1134 TLKEEQNKQKQEVNDLLQNEQELMKVMKTKHECONLESEPIRNSVKRESERESNOCNFKPQ 1193
QY 990 -----EVS-----RNLMHMENTGETKDEFOQKVVGDIKQDLEAKNT 1026
Db 1194 MLEVKEISLDSVNAQLVQLEAMLRNKLQJESKE--KECLOHELOTI--RGDLTSLN 1250
QY 1027 QTLTAD---VKDNETIEQOKIF-----SLIQEKNELQQMLESV 1062
Db 1251 QMQSQESIGLKDCEIDAEKYISGPHELSTSQNDNAHLQCSLQTTMKNLEKEICEIL 1310
QY 1063 IAEKEQLKTDLKEN-----EMTIENTQBELLLGDELKKQOEIVAQEKNHAIKE-----1112

Db 1311 QAEKYVELTELNDRSRSECITATRKWABEVGKLNLEVK-----ILNDSGLLHGLVEDIP 1365
QY 1113 -GELSRTCDR-----LAEVBEK-----LKEKSQQLQ--EKQOQLLNQVEE-----1149
Db 1366 GGEFEGEPNEQHPVSLAPLDESNSYEHLLTSDKEVQMHFAEQELKFLSLQSEHKILHQH 1425
QY 1150 ---MSEWQKKTINEIE-----NLKN---ELKNKELTLEH-----1176
Db 1426 QMSSKMSSELOTYYVDSLKAENLVLSNLRNFQGLVKEMQGLGLEGLVPSLSSSCVCPDSS 1485
QY 1177 ---AQKL-----METERLEL 1184
Db 1486 SLSSLGSSSFYRALLEOTGDMSSLLSNLEGAVSANQCSVDEFCSSLYQYVDSLKAENLV 1545
QY 1185 AQKL-----1188
Db 1546 STNLNRFQGLVKEMQGLGLEGLVPSLSSSCVCPDSSSLSSLGSSFYRALLEOTGDMSSLL 1605
QY 1189 ---NENYEEVKSITKERVKVLKELQKSFETERDHLRGYIRE 1225
Db 1606 SNLEGVVSANQCSVDEFCSSLOEENLTRKETPSAPAKGVBELESCEVYQSLKEKLEK 1665
QY 1226 IEATGLQTKBELKTAHILKHEOETIDELRR-----SVSE 1260
Db 1666 MESQIMKNKEIQEQLLSERQELDLCKRYLSENEQWQOKLTSVTLMEKSLAAEKK 1725
QY 1261 KTAQT-----INTQ-----DLEKSHTKLOEETPVUHE 1287
Db 1726 QTEQLSLEVARLOGLDLSSRSLGIDTADATQGRNESCDSIKESHTSETTERTPKHD 1785
QY 1288 ---EQELLPNVKVKSET-----1301
Db 1786 VHQICDKDAQODLNDIDIEKITETGAVKPTGBCSGEQSPDTNYPPGEDTKQGSSECSIEL 1845
QY 1302 ---QETMNELELLTQOSTTKDSTTL-----A 1324
Db 1846 SFGSNALVPMDFLGNQEDIHNLQVRVETSNENRLLHVIEDRDRKVESLNLNEMKELDS 1905
QY 1325 RIEMERLRLNKFO---ESOEIISLTKBERDNLAKTIKEALEVKHOL-----1368
Db 1906 KLHLQEVOLWKIEACIELEKIVGELKKENSLSLEKLEVFSCDHQELLQVRVETSEGLSD 1965
QY 1369 ---KEHIRTFLAKIQESQKQOSLANMKEKNETTKIVSEMEOFKPKDQALLRI 1419
Db 1966 LEMHADKSSREDIGNVAKVNDSW--KERFLDV---ENELSRIRSEKASIE--HEAAYLBA 2019
QY 1420 BTEMGLSKRLQESHEDEMS-----VAKEKDDQLQRLQVLOSES---DQLKEN 1464
Db 2020 DLEVQVTEKLCLEKNENKQKVVICLSELSVVTTSERQQLRGELDTMSKTTALDQLEK 2079
QY 1465 IKEIVAKHLETEEE---LKVAHCLKQEOETINELRVLNLSKEKETEISTIOKOLEAIND 1519
Db 2080 MKE-KTQELESHQSECLHCIOVAEABVEKTELLQTLSSDVSELLAKDKTHLOKQLEK 2138
QY 1520 ---KLONKIOEYKEEQNLKIQISEVQENVNVELQFKEHRAKOS--ALQSIIE 1568
Db 2139 DSQALSCLKELENOIAQL--NKEKELLVKSESLOARUSE---SDYEKNVSKALEAL 2193
QY 1569 SKWLEITNRLQESQEIQIMIK-----EKEEMKRVQOEALQI-----ERDOLKENTK 1614
Db 2194 VKGGEFALRLSTQEVHQLRGIEKLVRVIEADEKKQLHIAEKLERENDSLDKDVE 2253
QY 1615 EIVAKMKESQEKYOFKMTAVNETQEKRCRIEHLKQEPETO---KLNENIETEN 1667
Db 2254 NLERELQMESENG---ELVILDAENSKADEVETLTKTOIEEMARSLKIFELDLVTIRSEK 2308
QY 1668 IRLTQILHEN-----LEEMSVTKERD--DLRSVEET---LKVREDOLKENIRET 1712
Db 2309 ENLTQIOIQEQOQLSDELKLLSSFSLLEKKEQAEIQIKESKTAVENTLQNLQKE--LNEA 2367
QY 1713 ITRDLEKQBELKIVMHML-----KEHQ--ETIDKLRG--IVSEKTNFISNNQ--KOLEHSN 1762

Db 2368 VAALCGDQEIIMKATEQSDPPPIEBEHQLRNIEKIRARLEADEKQKQCVLQQLKSEBHA 2427
Qy 1763 DALKAQDLKIQEBELRIAHMLKEQETIDKLGIVSEKTDKLSNMQKDLNSNAKLQEKI 1822
Db 2428 DLKGRVENLERELTA---RTNQE-----HALEAENSKEVEETLKAKIEGMT 2473
Qy 1823 QELKANEHQILITUKD---VNETQK-----VSEMEQKQKQIQDQSL 1861
Db 2474 QSLRGLELDVVTIRSEKENLTNLOKEQERISELEIINSFENILOKEQEKQVQMKES 2533
Qy 1862 TSLKL---ETENLN-LAQELHENLEEMKSVKERNLNRVEETIKLRDQKESLQTKA 1917
Db 2534 TAMELQTLQKELNERNVAALNDQEA---KEQNLSSQVECELEKALQLOGLDEAKN 2590
Qy 1918 RDLIEIQELKTARMLSKHEKTDVKLRKISIKTIQISDIQKDLKSKD----- 1966
Db 2591 NYIVLOSSVGLLQFVEDGKQLEKDEEISRLKQIQDQEQVLVSKLSQVGEHQWLKEQ 2650
Qy 1967 -----ELQKIQELQKQELQLLRVKEDVNMHKKI-NEMEQKQKQEPNYLCKEM 2016
Db 2651 NLELRNLTVLEQKIQVLOSKNASLQDTLEVLOSSYKYLENELLT- 2698
Qy 2017 DNFQTLKHLSELEIRIVAKERDELRRYKESLMDRQDQFIATLRMIARDRONHQVKE 2076
Db 2699 DRMSFVEKVN-----KMTAKETELQREHMA-----QKTABLOE-----ELSGE 2738
Qy 2077 KRLLSDGQOHLMSLRKCSRIKELLKRYSEMDDHYECLNRLSLDLKEIEFHRIMKCLK 2136
Db 2739 KNLGELQULLLEIKSSKQDLKELTLENSLKSSJDCMHK--DOVEKE----- 2785
Qy 2137 YVLSYVTKIKGEHQECINKFMDFIDEVEKQELIKIHLQODCDVPS-RELRDLKLNQ 2195
Db 2786 -----GKVREE-----IAEYQLR-LHEAEKHQALLDNTNKQVEIQTRE-----KLIS 2830
Qy 2196 NMDLHTEELKQFSESEFFSIKTEFOVLNRKEMTOFLEWLNTRPDIKLNKGTOKEN 2255
Db 2831 K-----ECLSS--QKLEIDLKSSKSELNLSKATTTQILEELKTKMDNLKYVNLKCN 2884
Qy 2256 DRICQVNNFFNNRIIIMNESTEFERSATISKEWQDLKSLKEKQKLPKNVQTLKSL 2315
Db 2885 ERA-----QKMKLLIKSQLEKEKELLQKLSQ-LQAAQEK----- 2921
Qy 2316 ASGAQNPPTQDNKNPHVTSRATQLTKIRIENLSLHEAKESA---MHKESKIKMQKE 2372
Db 2922 ----QKTGTWMDTK-----VDELTE-IXELKETLEKTEADRYLKYCSLLISHEK 2969
Qy 2373 LEVTNDIILAKQAVHESKCLKTKETIQ--VLQKVAGAKPYKEEIED-LKMKLVKI 2429
Db 2970 LEK-----AKEMLETOVHLCSQOQSDRGSPFLGVPVPGSPIPSVTEKRLSSQNK 3024
Qy 2430 DLEKMKNAKEFEKEISATKATVE-YOKEVIRILLRENLRSQQAQDTSVISEHTDPOPSNK 2488
Db 3025 SKQRQSSGIWENGRTPTATPESPKSKKAVMSGI---HPAEDT-----EGTEPEPGL 3077
Qy 2489 PUTCGGGG-----IVONTKALILKSEHRLK-EISKLKQONBQLIKQKNELLS 2537
Db 3078 PEVVKKGFPADIPGKTSPYILRTTWMATRTSPRLAAQKLALSPLSIGKENLAESSKPTAG 3137
Qy 2538 NNQHLNENKWTERTKREAHQVTCENSPKSPKVTGTAGKKQITPQCKERNLQDPV 2597
Db 3138 GSR---SQKVKV-----AQSPVDSGTILREP---TTKSVFVNNLPERS 3175
Qy 2598 PKESPSCFFDGRSKSLSPHPVRVFDNSSLGLCPVQVAGAESVDSQGPWHAAS 2653
Db 3176 PTDSREGRLVRGRILVPAPK-----LDWSQLAV-RTVRSSEALCVSD---PWEVOS 3223

RESULT 8

PCT-US95-16216-1

; Sequence 1, Application PC/TUS9516216

; GENERAL INFORMATION:

; APPLICANT: Yen, Timothy J.

; APPLICANT: Rattner, Jerome B.

; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
; TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patencin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/16216
; FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA: US 08/353,700

APPLICATION NUMBER: 09-DEC-1995

ATTORNEY/AGENT INFORMATION:

NAME: Reed, Janet E.

REGISTRATION NUMBER: 36,252

TELEPHONE: (215) 563-4100

TELEFAX: (215) 563-4044

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 3248 amino acids

TYPE: amino acid

STRANDEDNESS: not relevant

TOPOLOGY: not relevant

MOLECULE TYPE: protein

HYPOTHETICAL: NO

ANTI-SENSE: NO

PCT-US95-16216-1

Query Match 7.8%; Score 1040.5; DB 5; Length 3248;
Best Local Similarity 19.6%; Pred. No. 6.7e-41;
Matches 635; Conservative 583; Mismatches 1033; Indels 985; Gaps 132;

Qy 154 IREDVNRVYVADLVEEVYTTSEMALKWITKEKSHYGETK-----MNQSRSRSHYTFM 209
Db 237 IRRDFSASYFGELE---VTPSRSTLIQIGKRDANSFFGNSSPHLLDQLKQAOELRNK 293
Qy 210 I--LESREKGEPSNCEGSV-KVSHLNL-----VDLAGSER-----AAQTGA 248
Db 294 INELRLQGHKEKMGQVNFQELQLEKAKVLEIEKVLNCKRDELVRTTAQDOA 353
Qy 249 GYRLKEGNNINRSLFILGQVVKLSDGQVGGFINYRDSKLTRILQNSLGNPKTRICTI 308
Db 354 STKYT-----ALEQKAKLTEDLSQONAESARCS--LEQKIKEKEK----- 394
Qy 309 TPVSFDETLTALQ-----FASTAKMNTPTVNEVSTDE--ALLKRYRK 350
Db 395 ---EFOEELSRQORSFQTLDOECIQMKARLTQELQAKNMHNVLAELDKLTSVKQOLEN 451
Qy 351 ETMDLKKOLESVLETRAQAMEKQOLAOLLE---EKDLQKVQNEK-----TEN 396
Db 452 NLEEFKQKLCAEQAFQASQIKENELRRMEEMKENLLKSHSQKAREVCHLEAEKCN 511
Qy 397 LTRMLVTSSTLTQOELKAKRRVTVCLG---KINKMNSNYAQDFNPTNTTTTKTKL 453
Db 512 IKQCL--NQSQNFAEMKAKNTSQETMLRDLQEKINQOENSILTLEKLAVADLEKQDC 569
Qy 454 SINLLREIDESVCSSESVFNT---LDTLSEIE-----WNPAT-KL 490
Db 570 SQDLLKQREHIEQLNDKLSKTEKSKALLSALEKKEEYELKEEKLTFSCWKSENEKL 629
Qy 491 LNQ-----ENTESELNSLRA-----DYDNLVLDYEQLRTEKEEMELKLEKNDLD 535

Db 630 LTOMSEKENLQSKINHELTCLKTQOIKSHYNERV---RTLEMDRENLSVEIRNLHVL 686
Qy 536 EFEALERKTK--DOEMQLIHEISNLKMLVKHREVYNQDLENELSSKVLELREXDOIKK 593
Db 687 DSKSVETOKLATMELQOAKFSDQK-----HOKEIENMCLKTSQLTQGVDELEHK 738
Qy 594 LQ-----EYIDSQKLENTKMDLSYLESIEDPKOMKOTLFDATV-----ALDAK- 638
Db 739 LQLLSNEIMDK--DRCVQDLHAEVESLRDLKSK---DASLVNEDHQSRLLAFAQDP 791
Qy 639 -----RESAFLASENLELKEKMKELATTKOMENDIQLQVSOQLBAKKMQVDLEK 688
Db 792 AMHHSFANIIGEOGSMPSSECRLEADQSPKSAILLQNRVDSLEFSLSESQOMNSDLQK 851
Qy 689 ELQSAFNEITKLSIDGKVPKDLL-----CNLELEGKITDLOKEL-----NKEVEEN 736
Db 852 QCE-----ELVQ-----IKGEIBENLMKAROMQSFVAETSQISKLOEDTSAHQNVVAETL 903
Qy 737 EAL-----REEVILLASLSPSEVERLRKE---IQDKSEELHIITS-----BKDLKFSEV 785
Db 904 SALENKKEKQLLNDKVETEQAIBQELAKSNHLLSELSKELQLLSETLSLEKEMSSIIS 963
Qy 786 HKRSRVOGLLEBRIGHTKD--DLATQSNYSKSTQDFONFKTLHMDFKQKVMVLENER 842
Db 964 LMKREIEELTOENGTLEKINASLNQEKWNLIQKSESFANY-----IDERKSGISELSDQ 1017
Qy 843 MNOEIVNLSKEAQKFDSSIGALKTEL SVKTOELQKTRV-----QERLNEMEQ 891
Db 1018 YKOEKLIILQRCSE---TGNYEDLSQYKAAQEKNSKLECLINECTSLCENRKNLEQ 1073
Qy 892 LKEQLENRDSPIOT-----VEREKTLI---TEKLOTLLEVKTLTO-----EKDDLKQLQE 939
Db 1074 LKEAPAKEHQEFTKLAFAEERNQNLMELETVQALRSEMTDNQNSKSEAGLQKQIM 1133
Qy 940 SLQIERDQKSDIHTVANNIITQOLRNALLESKHOETI--NTLAKSLSE-----989
Db 1134 TLKEEQNKQKVEVNDLQENQELMKVMKTKHECONLESEPIRNSKRESERNQCNFKPQ 1193
Qy 990 -----EVS-----RNLHMENTGETKDEPQKQMVGIDKKQDLKAKNT 1026
Db 1194 MDLEVKESLDSYNALQVLEAMLRNKNELKLOSEKE--KECLOHELOTI--RGDLTSLNL 1250
Qy 1027 QTLTAD-----VKDNHIEQQRKIP-----SLIOEKNELOQMLESV 1062
Db 1251 QDMQSQEISGLKDCIDAEBKYISGPHSLSTQNDNAHLQCSLOTTMKNLEKICEIL 1310
Qy 1063 IAEKQKLTDLKEN-----IEMTIENQOELRLGDELKQOEBIVAQEKHAIKKE-----1112
Db 1311 QAEKYELVTELNDRSSECTATRKQABEVGKLLNEVK-----ILNDDSGLLHGEIVEDIP 1365
Qy 1113 -GELSRTCDR-----LAEBEEK-----LKEKSQOLO--EKQOQLLNVQEE-----1149
Db 1366 GGEFGEQPNQHPVSLAPIDESNYSYHETLSDEKQVHFAELOEKFLSLQSEHKILHDQH 1425
Qy 1150 -----MSBMQKINBIE-----NLKN-----ELANKELTLEH-----1176
Db 1426 CQMSSKMSSELOTVVDSLKAENLVLSNLRNFQDVLKEMQGLGEBGLVPSLSSSCVPDSS 1485
Qy 1177 -----METERLEL 1184
Db 1486 SLSSIGDSSFYRALLEQTGMSLLSNLEGAVSANQCSVDFVCSLSQTVVDSLKAENLV 1545
Qy 1185 AQKL-----1188
Db 1546 STNLRNFQDVLKEMQGLGEBGLVPSLSSSCVPDSSSLSGDSSFYRALLEQTGMSLL 1605
Qy 1189 -----NENYEVKSTIKERKVLKELQKSFETERDHLRYIRE 1225
Db 1606 SNLEGVVSANQCSVDFVCSLSQENLTRKETPSAPAKGVEELESLEYVRSLEKLEEK 1665
Qy 1226 IEATGLOTKEELIAHILKHOETIDELR-----SVSE 1260
Db 1666 MESQIGMKNKBIQELQELSSERQELDLKQKYLSENEQWQKLTSTVTLEMESKLAEEK 1725

Qy 1261 KTAQI-----INTO-----DLEKSHTKLOEIPVLHE 1287
Db 1726 QTEQLSLEVARLOQLDLSRSLGIDTDAIQGRNESCDISKEHTSETTERTPKHD 1785
Qy 1288 -----BOELLPNVKVYSET-----1301
Db 1786 VHQICDKDAQOOLNIDIEKITEGAVKPTGECSEQSPDNYEPFGEDKTQGSSECISEL 1845
Qy 1302 -----QETMNELELLTEQSTTKDSTTL-----A 1324
Db 1846 SFGSPNALVPMDFLGNQED IHNQLRVKETSNEENLRLLHVIEDRDRKVESLINEMKELDS 1905
Qy 1325 RIEMERLRLNKFO---ESQEEIKSLTKERNLTKITKEALEVKHQD-----1368
Db 1906 KHLHQVQLATKIEACIELEKIVGELKENSDDLSEKLEFYFSCDQHELLQORVETSEGLNSD 1965
Qy 1369 -----KEHIRETLAKIOSQSKOBOSLNMMKEDNETTKIVSEMEQFKPKDSALLRI 1419
Db 1966 LEMHAKSSREDIGDNVAKVNSW--KERFLDV---ENELSRIRSEKASIE--HEALYLEA 2019
Qy 1420 EIMLGLSKRLQESHDEMS-----VAKEDDLQRLQOEVLOSES---DOLKEN 1464
Db 2020 DLEVVOETEKLCLEKONENKQVIVCLEEELS VVTISERNQLRGELDTMSKTTALDQSEK 2079
Qy 1465 IGEIVAKHLETEEE---LKVAHCCLKEQEBETINELRVNLSEKETEISTIQKOLEAND 1519
Db 2080 MKE--KTOELSHQSECLHCIOVAEAEVKETELLOTTSSDVSELKDKTHLOEKLOQSEK 2138
Qy 1520 -----KLONKIOEIVEKEQNLNKOI SEVOENNELKQFKHEKAKDS--ALOSIE 1568
Db 2139 DSQALSUTKCELENOIAGL--NKEKELLAVKESLOARLSE---SDYEKLVNSKALEAAL 2193
Qy 1569 SKMLETNRLQESQEBEIQIMIK-----EKEMKRVQOALQI-----ERDQKENTK 1614
Db 2194 VEKGEFALRLSSTQEBEHQLRGTEKLRVRIEADKQHLHAEKLRERENDSLDKOVE 2253
Qy 1615 EIVAKMKSQEKVQFKMTAVNETQKMCLEHLKEQFETQ-----KLNLENIETEN 1667
Db 2254 NLERELOWSEBQ-----ELVILDAENSKAEVETLUKTOEAMARSILKIFELDLVTRSEK 2308
Qy 1668 IRLTOLIHEN-----LEEMRSVTKERD--DLRSVEET---LKVERDOLKENLRET 1712
Db 2309 ENLTQIQOEBQKQOLSELKLLSSFKSLLEKEQAIBQIKESKTAVEMLQNLKE--LNEA 2367
Qy 1713 ITRDLEKQEBELKIVMHML-----KSHQ--ETIDKURG--IVSEKTHNISNQ--KOLEHN 1762
Db 2368 VAALCGDQEIIMKATQSLDPPTEEBHQLRNSIEKLRARLEADEKKQLCVLOQLKSEHHA 2427
Qy 1763 DALKAQDLKIQEBELIAHMHLEKQOETIDKLRGIVSEKTDKLSNMQKDLSENSAKLOEKI 1822
Db 2428 DLLKGRVENLERELEIA---RTNOE-----HAALAEANSKGEVETLKAKLEGMT 2473
Qy 1823 QELKANEHOLIITLKKD---VNETOKK-----VSEMQLKQIKQDOSL 1861
Db 2474 QSLRGELDLVTVIRSEKENLTLNLOEQEIRISELEIINSSFENILQEKQEKVQMKKSS 2533
Qy 1862 TLKSL---BIENIN--LAQELHENLEBMKSVMKERNLRRVETLKLRLQELKESLOETKA 1917
Db 2534 TAMEMLQTLKELNERVAALHNDQEAACKA---KEQNLSQVECELEKALQLGLDEAKN 2590
Qy 1918 RDELETOELKTARMLSKHEKTEVTKLRKIKSEKTIQISDIOKDLDKSKD-----1966
Db 2591 NYVILQSSVKGIIQEVEDQKLEKQDEBESILKQIQOQEQVSKLSQVEGHEHQLWKEQ 2650
Qy 1967 -----ELQKKIQELQKELQLIRVKEDVNMSSHKI--NEMEQLKQKQFEPNYLCKCEM 2016
Db 2651 NLELRNLITVELBQKIOVLQSKNASLODTLEVLQSSYKKNLENELETK-----M 2698
Qy 2017 DNFQITKKLHESLETRIVAKERDELRIRKESIKMERDQFIATIRMIARDRONHQVKPE 2076
Db 2699 DKMSFVEKVN-----KMTAKETELQREHMA-----OKTAELOE-----ELSGE 2738

```

QY 2077 KRLSDGQOHHMESLREKCSRIKELLKRYSEMDHYECLNRLSLDLKLEKHEFHRRMKKLK 2136
Db 2739 KNRLAGELQLLEBKSKDLQKBLTLENSKSLDCHK--DQVEKE-----2785
QY 2137 YVLSYVTKIEEQHECINKEMDPIDVEKQKELLIKIHOHQDCVPS-RELRLKLNQ 2195
Db 2786 -----GKVEE-----TAEQLR-LHEAKKHQALLDTNKKQVEVIQTYRE----KUTS 2830
QY 2196 NMDLHIEELKDFSEFPFSIKTEFOQVLNRKEMTQFLREWLNTRFDIEKLKNGIKEN 2255
Db 2831 K-----ECLSS-QKLEIDLKSKSEELNLSKATQILLELKTKKMDNLYVNLQLEN 2884
QY 2256 DRICQVNNFNRIIAMNTEPERSATISKWEQDLKSKNEKLFKNYQTLKTSL 2315
Db 2885 ERA-----QGRMKLLIKSKQLEERKEILQKELSQ-LQAQEK-----2921
QY 2316 ASGAQVNPPTQDNKNPHVTSRAQTLTTEKIRELENSLHEAKESA---MHKESKILMKQE 2372
Db 2922 ----QKTGVWDYK-----VDELITTE-IKELKTELEKTKADEYLDKYCSLLISHEK 2969
QY 2373 LEVNDIIAKLQAKVHESNKKLEKTKETIQ--VLQDKVALGAKPYKEEIED-LKMKLVKI 2429
Db 2970 LEK-----AKEMLETVAHLCSQKDSRGSPILGPVPVPGSPISVTEKRLSSGONKA 3024
QY 2430 DLEKMKAKEFEKEIGATATVB-YOKEVIRLLRENLRRSQOQDTSVISEHTDPOPSNK 2488
Db 3025 SGKRQRSSGIWENGRGPTPATPESFSSKSKKAVMSGI---HPAEDT-----EGTEFEPEGL 3077
QY 2489 PLTCGGGSG-----IVQNTKALLIKSEHIREK-EISKLQKQNEOLIKQKNELLS 2537
Db 3078 PEVWKGFPADIPGKTSYPILRRTWATTSPLRLAAQKLSPLSLGKENLAESSKPTAG 3137
QY 2538 NNQHLNENYKTKERTLKREAHQKVCENSPKSPKVTGTGASKKKQITPSCCKERNLODPV 2597
Db 3138 GSR--SQVKV-----AQRFVDSGIIIRP--TTKSVFVNNNIPERS 3175
QY 2598 PKESPSCFFDSRSKSLPSPHPVRYFDNSLSGLCPVQVAGAESVDSQGPMPHASS 2653
Db 3176 PTDSPREGLRVGRGLVLPAPK-----LDWSQLAV-RTVRSEALCVSD---PWEVQS 3223

```

RESULT 9

```

US-08-328-254-6
; Sequence 6, Application US/08328254
; Patent No. 5710022
; GENERAL INFORMATION:
; APPLICANT: Zhu, Xueliang
; APPLICANT: Lee, Wen-Hwa
; TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein
; CORRESPONDENCE ADDRESS:
; NUMBER OF SEQUENCES: 8
; ADDRESSER: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/328,254
; FILING DATE: 24-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/141,239
; FILING DATE: 22-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-CJ 1191

```

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2482 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-328-254-6

```

```

Query Match 7.8% Score 1037; DB 1; Length 2482;
Best Local Similarity 20.9%; Pred. No. 7.3e-41;
Matches 580; Conservative 508; Mismatches 951; Indels 734; Gaps 116;

```

```

QY 320 LQFASTAKYMNTPPYNEVSTDEALLKRYKEIMDLKQLEEVSLSETRAQAMEKDQLAQL 379
Db 15 LETCLATQIKSHEYNERVRLTEMDRENLSVEIRNLHNVLDSKSVETQKLAWELOQK 74
QY 380 LEEKDLLQKVQNEKINLRLMLTVSSSLTQ-----OELKAKR 417
Db 75 AEFSD--QKHQKE-IEN--MCLKTSQLTQGVEDLEHKLQLLSNEIMDKDRCYQDLHAHY 128
QY 418 KRVTWCLGKINKMKNVADQFNIPNTITTKTKLSI-----NLLREIDE 463
Db 129 E-----SLRDLKSKDASL-----VNEHDQSLLAFAQDQOPAMHHSFANIIGE-- 171
QY 464 SVCSSESVFNTLDTLSEIENPATKLINQENIES-----ELNS-LRADYDNLVLVD 513
Db 172 ----QGSMPSESECELEADQSPKNSAILQNRVDSLEFSLESQKNSDLQKCEBLV-- 225
QY 514 YEQLRTEKEEMELKLEKNDLDEFEALERTKTKQDEMQ-----LIHEISNLKMLVHREV 568
Db 226 --QIKGEIENLMKAFQMSQSFVAETSQRISKLEQEDTSAHQNVVAETLSALENKKELOL 283
QY 569 YNQDLNENLSSKVELLREK---EDQIKLQEVYDSQKLENIKMD--LSYSLESIEDPKOM 623
Db 284 LNDKVETEQAETQELKKSNNHLLSDLSKELQLLSLSELEKKEKSSIIISUNKREIBELTQE 343
QY 624 KOTLFDATVALDAKRESAFLRSENLEKKEKMLATYKQWENDIQLYQSQLEAKKKMQ 683
Db 344 NGTL-----KEINASLNQKWNLLQKSESFANYIDEREKSISELSDQYKQEKIL 393
QY 684 VDLEKELQSAFNEITYKLSLDGKVPKOLLNLEEGKITDLOKELNKEVNEALREEV 743
Db 394 LQRCBETGNAYE-----DLSQKYKAAQAEKSKLE---422
QY 744 ILLSELKSLPSEVERLRKEIODKSEELHIITSEKDKLPSEVVHKSRSVQGLAEIIGKT-- 801
Db 423 CLLNCTSL---CENRKELEQLKE---AFAKEHQEFITKLAFABERNQNLMELETVOQ 476
QY 802 --KDLATTSQNYKS-----TDQEFQNFKTLHMDFEQKYMVLSEERNMQEIVNLSKEAQ 855
Db 477 ALRSEMTDNQNNKSEAGGLQKEIMTLKEEQNMQKEVNDLLQENEQL-MKVMKTKHECQ 535
QY 856 KFS-----SLGALKTELSYKTO-ELQEKTRVQERLNMEMQOLKQLENRDSPLQTV 906
Db 536 NLESEPIRNSVKRESERNQCNFKPQMDLEVKEISLDSYNAQLVQLEAMLRNKLKQLQES 595
QY 907 EREKTLITEKLOQTLEEVKT-----LTQEKDCLKQLQ-----ESLQIERD 946
Db 596 EKEKELQHELOTTRGDLTSLNMQMSQSEISGLKDCSDAEKYSIGFHEHLSSTQNDNA 655
QY 947 QLKSDIHDTVMNMDTQQLRNASLESKQHOET-INTLKS-----KISEVSNLH-- 996
Db 656 HLQCSLQTTMN-KLNELEKICEILQAEKYVELTELNDSRSECITATRWAEVGGKLLNEV 714
QY 997 --MEENTGETKDFQQRKMGV-----IDKKQDLEAKNTQTITADVKONE 1037
Db 715 KILNDDSGLLHGLVEDIPGFGFQEPNQHPVSLAPLD-----ESNSVEHITLSDKEVQ 769
QY 1038 I--TEQQRKIESLQOE-----KNELQOMLESVIAEKQLKT-----DLK 1074

```


Db 770 MHFAELQKFLSLQSEHKILHDQHCSSKMSSELQTYVDSLKAEINLVSLNRFQGDV 829
QY 1075 ENITEMIE-----NOBELLLGDELKKQOEIIVAOEKHAI--KKEGELST- 1118
Db 830 KEMQLGLEGLVFLSSSCVPSDSSSLGDS-SFYRALLETQTDMSLLNLEGAVSANQ 888
QY 1119 -----CDRLAEVEELKKEKSOQLQKQOQLNVQEMSMQKINEIENLKNELNKE 1171
Db 889 CSVDVEFCSSLOENLFRKTPSPAPAKGVELESICEVYRQSLKLEKEMESQIGMKNKE 948
QY 1172 L-TLEH-METERLEL-----AQKNENYEVVKSITKERVUKVLEKQSFETERDHLRGVIRE 1225
Db 949 IQLEQLLSERQELDCRLQYLSENEQWQOKLT---SVTILEMSKLAEEK-----KQ 998
QY 1226 IEATGLQTKBELKTAHHLKHEQETIDELRSV-----SEKTAQIINTQ-DLEKSHTKLOE 1280
Db 999 TEQLSL-----ELEVRLQL-----QGLDLSRSLGIDTEDAIQGNESCDSIKETSETT 1050
QY 1281 EIPVLHE-----EQLLPNVKVKSET----- 1301
Db 1051 ERTFKHDVHOICDKDAQODNLNDIEKITETGALKPTGSCSGEQSPDNTNVEPPGBDKTQGS 1110
QY 1302 -----QETMNEBELLEQSTTKOSTTIL----- 1323
Db 1111 SECISELSPGPNALVPMDFLGNQEDIHNLQVRKETSNEMLRLHVIEDRDRKVESLLN 1170
QY 1324 -----ARIENRRLNKEFQ-----ESQEEIKSLTKERDNLKTIKEALEVKHDOL----- 1368
Db 1171 EMKELDSKHLQEVOLMTKIBACHELEKIVGELKENSDDLSEKLEYFCSDHQLLQVET 1230
QY 1369 -----KEHIRETLAKIQESQKOSQOSLNKMEKONETTKIVSEMBOFPKP 1412
Db 1231 SEGNSDLEMAHDKSSREDIGDNVAKVNSW--KERFLDV--ENELSRIRSEKASIE-H 1284
QY 1413 DSALLRIBIEMGLSKRLQORSHDEMS-----VAKEKDDDLQRLQEVLOSES-- 1458
Db 1285 EALYLEADLEVQTEKLCLEKONENKQVIVCLEBELSVTTSERNQELGELTMSKTTA 1344
QY 1459 -DOLKENIKETIVAKHLEFEE-----LKVAHCLAKEOETINELRVNLSEKETETSTIQK 1512
Db 1345 LDQSEKMKK-KTOELSHQSELCIQAIAEVEKTELTQTLSSDVSSELLKDKTHLOE 1403
QY 1513 QLEAIND-----KLONQIBIYEKEBOLNIKOISEVOENVELKQPKFHEHRKAKDS 1562
Db 1404 KLOSLKESQOALSFLKCELENQIQL-NKEKELLVKESSLOARLSE-----SDYEKLNV 1458
QY 1563 -ALQIESKMLELTVRLQESQEIQIMIK-----EKHEKRVQERALQI-----ERD 1607
Db 1459 KALEAALVEKGBFALRLSSTOBEVHQLRRGIEKLVRIEADKKQLHIAEKLKEREREND 1518
QY 1608 QLKENTKEITVAKWKEQKEYQFLKMTAVNETQEKMCETIHLKEQPETQ-----KML 1660
Db 1519 SLKDKVENLERELQSEHQ-----ELVILDAENSKAEVETLKTQIEEMARSLKVPFELD 1573
QY 1661 ENIETENIRLQILHEN-----LEEMRSVTYKRD--DLRSVEET--LKVERDOL 1705
Db 1574 VTLRSEKENLTQIQEKQQLSELDKLSSFKSLLEEKEQAEIQIKESBKTAVEMLQNLQ 1633
QY 1706 KENLRBTITRDLKQBELKIVEMHL-----KEHQ--ETIDKLRG-IVSEKTNESNMQ-- 1755
Db 1634 KE-LNEAVALCGDQIMKATEQSLDPPTEEBEHQLRNSIEKLRLEADEKXQCVIQL 1692
QY 1756 KOLEHSDALKAQDLKQIERLTAHMLKEQOBTIDKLRGIVSEKTDKLSNMKOLENSN 1815
Db 1693 KSEHHAHLKGVENLERELETA-----RTNQE-----HALEAENSKGEVETLK 1738
QY 1816 AKLQEKIQELKANEHOLITLKKD-----VNETQKK-----VSEMQLKK 1854
Db 1739 AKTEGQTSLRGELDVVITRSEKENLTNELQKEQERISELITINSFENILOKEQEKV 1798
QY 1855 QIKDQSLTSLSKL---EINIEN-LAQELHENLEEMKSMKERNLRRVEETLKLEROQKE 1910
Db 1799 QMKEKSTAMEMLQTLKELNERNVAALHNDQEAACKA---KEONLSQVECLELEKALQLLQ 1855

QY 1911 SLOETVARDLEIQOELKTARMLSKHEKETVDKLRKISEKTIQISDIQKDLKSKD--- 1966
Db 1856 GLDEAKNNYIVLQSSVNGLIQVEDDGKQLEKDEEISRLKNQIQOQOVLVSKLSQVEGE 1915
QY 1967 -----ELOKKIQLOKQELQLLRVKEDVNMHKKI--NEMEQLKQKQEPN 2009
Db 1916 HQJWKEQNLERNLTVLEQKIQVQKSNASLODTLEVLQSSYKYLENELELT- 1969
QY 2010 YLCKCENDNFQTKLHESLEERIVAKERDELRIKESLKMERDQFIATULREMIARDQ 2069
Db 1970 -----NDKMSFVEKVN-----KMTAKETELQEMHEMA-----OKTAELOE----- 2005
QY 2070 NHQVPEKRLSDQOQHLMESLREKSRIKELLAKRYSEMDDHYECLARLSLDLKEBHF 2129
Db 2006 --ELSGEKNRLAGELQLLEELIKSSKQDLBELTLENSELKSLDCMKA--DOVEKE----- 2057
QY 2130 RIMKKLVLSVYTKIEEQHECINKPEMFDIVBEKQKELLIKIQLHQQDQVPS-REL 2188
Db 2058 -----QKVREE-----IAEQLR-LHEAKKHQALLDNTKNQYVEIQTIRE- 2098
QY 2189 RDLKLNQMDLHIEILLKDFSEFPSPKTFEQOVLNKRKEMTQFLEEWLNTRFDIEK 2248
Db 2099 --KLTSK-----EECLSS-QKLEIDLKSSKBEINLSLKATQILEELKTKTMDNLKV 2149
QY 2249 NGIOKENDRICQVNNFFNNRIIAMNESTEPEERSATISKEMEQDLKSLKEKNEKLPKNY 2308
Db 2150 NQLAKENERA-----QGMKLLIKSCQKQLEKEILQKELSQ-LQAAQEK----- 2193
QY 2309 QTLKATSLASGAQVNPPTODNKNPHVTSRATQITTEKIRELENSLHEAKESA---MHKESK 2365
Db 2194 -----QKTGTVMDTK-----VDELTTE-IKELKETTLEKTEADEYLDKYCS 2234
QY 2366 IIRKQKELEVTNDIIAKLOAKVHESNKCLEKTKETIQ--VLQDKVALCAKPYKEIED-L 2422
Db 2235 LLISHKLEK-----AKEMLETQVAHLCSQKQSDRSRGLLGPVVPSPSPVTEKRL 2289
QY 2423 KMLKVIDLEKMKNAKEFEKISATKATVE-YQKEVIRLLARENRRSQOQDTSVISEHT 2481
Db 2290 SSGQNKASGRQSRSGIENGEGGPTPATPESFKSKKAVMSGI---HPAEDT-----EGT 2342
QY 2482 DPQPSNKPLTCGGSG-----IVQNTKAILKSEHIRLEK-EISKLQKQNEQLIK 2530
Db 2343 EFEPGLBEVVKGADIPDTGKTSYILRRITMATRTPRLAAQKLLSPISLGENLA 2402
QY 2531 QKNELLSNNQHLNNEVTKWERTKREAHKQVTCNSPKSPKVTGASKKKQITPSOCKE 2590
Db 2403 SSKPTAGGSR--SQKVQV-----AORSPVDSGTILREP--TTKSVPV 2440
QY 2591 RMLQDPVKESPKSCFFDSRSKSLSPHPVRYFYDNSSSLGLCPVQNAQESVDSQPGPW 2650
Db 2441 NNLPERSPDTSREGRLVRKGRIVPSP-----KAGLESKGS----- 2477
QY 2651 ASSGKDVPECKTQ 2663
Db 2478 -----NCKVQ 2482

RESULT 10

US-09-595-684B-23
; Sequence 23, Application US/09595684B
; Patent No. 6544766
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Ohashi, Cara
; APPLICANT: Sakowicz, Roman
; APPLICANT: Vaisberg, Eugeni
; APPLICANT: Wood, Kenneth
; APPLICANT: Yu, Ming
; TITLE OF INVENTION: Human kinesins and methods of producing
; TITLE OF INVENTION: and purifying human kinesins
; FILE REFERENCE: cytop036
; CURRENT APPLICATION NUMBER: US/09/595,684B


```

; CURRENT FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: 09/295,612
; PRIOR FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 1231
; TYPE: PRT
; ORGANISM: Human
; US-09-595-684B-23

Query Match
Best Local Similarity 7.1%; Score 946; DB 4; Length 1231;
Matches 343; Conservative 226; Mismatches 479; Indels 254; Gaps 43;

QY 7 VAVCVVRPLNSREESLG-ETAQVYKTDNNVYQVDSKSFNDRFVHGHNETTKNVEE 65
Db 10 VRVALRCRPLVPKEISEGQCMCLSFVPGEQVVGTD--KSFTYDFVDFPSTQEVEFNT 67
QY 66 IAAPILDSAIQYNGTIFAYGQTASGKTYTMG-----SEDLGVIPRAIHDFOKIK 118
Db 68 AVAPLIKGVFGYNATVLAQGTSGKTYSMGAYTAQENETVGVIPRVIQILFKED 127
QY 119 KPDPREFLRVSMEIYNETITDLCCTQKMLPIREDVNRNVYVADLTVEEVYVYSEMA 178
Db 128 KKSDFEFTLKVSLEYNEYNEEILDLLCPREKAQINIREDPKEGKIIVGLTEKTVLVALDT 187
QY 179 LKWITKESRHYGETKMNORSRSHITPRMILESREKGEPSNCEGSKVSHNLVLDLAG 238
Db 188 VSCLEQGNRSRTVASTAMNSQSSRSHTIFTI-----SLEQGGKSKDNSSFR-SKLHLVDLAG 243
QY 239 SERAQTGAAGVRLKEGNCNINSLFILGOVKKLSQGVGFYNYRDSKLTIRLQNSLGG 298
Db 244 SEROKTKAEGDRLKEGINLRGLCLGNVISALGDDKGGFAPYRDSKLTIRLQNSLGG 303
QY 299 NPKTRITICITPV--SFDETLTALQFATAKMNTPYNEVSTDEALLKRVKKEIMDLK 356
Db 304 NSHTLMACVSPADSNLEETLRLYADRARKIKNKPIVN-----IDPQTAELNHLK 355
QY 357 KOLEEVS-----ETPAQAMEKDQALAEELKDLQKQVNEKIENL 397
Db 356 QVQVQVQLVLLQAHGTLPGSTIVSEPNLSQMEKNQ--SLVEENEKLSRGLSEAGQT 413
QY 398 TRML---VTSSLTLOELKAKRRVTWC---LGKINMKNSYADQFNIPNTIKTH 451
Db 414 AQMLERLIWTEQANERKNALBELRCHAAKLDLQKLVELQELXENVEIICNLQOLLIT 473
QY 452 KLSINLLRDESVCSVDVFNLTDLSEIENWPATK-----LNNQENI 496
Db 474 QLS-----DETACMAAAIDTAVEQEAQVETSPETSRSSDAFTTQHARLQAQMSKELV 526
QY 497 E-----SELNSLRADY-DNL---VLDYQLRTEKEEMELK-KKN 532
Db 527 ELNKALALKEALARKMTQNDSQLQIQYQNIKEPELEVINLQKBEELVLELQATAK 586
QY 533 DLDEFEALERTKQDQEMQLIHEISNLKLVKHREYVNDLENELSSKVELLAREKEDQIK 592
Db 587 DANQAKLSERRKRILE- -QIADLKKL-----NEQSKLLKKESTERTVS 633
QY 593 KLOEVIDSQLENIKMDLSYSLESDPKOMQTLTDAETVAL---DAKRESAFIRSE-N 648
Db 634 KLNQEIIRMMKNQVQL-MRQMKDEAKFRQWQKR-DKEVIQLKERDRKQVELLKLERN 691
QY 649 LS-----LKEKMKELATTVKOMENDIQLVQSOLEAKKQVLELQSAFNEIKL 703
Db 692 FOKQSNVLRKTEBAANRUKDALQ-----KQREVAOKRKEOTOSRGMEFTA----- 739
QY 704 IDGKVPKOLLNLELGKITDLOKELNKEVEBEALREEVILLSELK-SLPEVERLRKE 762
Db 740 --ARVKNWLGNEIENVVSTEARHNLNLEDRKILAQDAQVQKESGENPPKLRRR 797
QY 763 IQDKSEELHIITSEKDKLFSEVHVHESRVQGLLEIGIKTKDLDLATQSNYKSTQDEQNF 822
;

```

```

798 TFSLTEVRGVSSESDSITTKQIESLETETEMFRSAQIADLOQLKLLDAESDR-PKQRWENI 856
QY 823 KTLHMFPEQYKVMVLENERMNOIRIVNLSKEAKFDSLSGALKTELKYQTOELQEKTR-- 880
Db 857 ATI---LEAKCAL-----KYLIGELVSSKIQVSKLESSLQSKTSKSCADMQLPFEERNHF 908
QY 881 ----EVOERLNEQLEQENRDSPLQVVEREKTILITEKLOQTLSEVKTILTEQKDDLK 935
Db 909 AEIETELQAEVLVRNEQ-----QHEKVLVLLSLOQSQMAE-----K 945
QY 936 QLOESLQIERDQLKSDIHDVTNNMNIQTQOLRNALSLKQHOETINTLKSKIS--EEVSR 993
Db 946 QLESVSSEKQQLLS-----TLKQDELEKQREVCEQOQLLRENEIILKLTLLQVSR 1001
QY 994 NLHMEWT-----GETKDEPOQKMGVI--DKKQDLEAKNTQTLTADVKD 1035
Db 1002 QKHLPKDTLLSPDSSEFYVQPKPSRVKKEKFLQESMDIEDLKCYSEHSVNEHEDGDD 1061
QY 1036 NEIIEQQRKIFSLI-----QEKNE--LQOMLESVIAKEQL 1069
Db 1062 DEGDDEWTKLVNVSRNKNIQCSCKGMCNKQCGCRKQKSDGVDCDDPTKCRNRQ 1121
QY 1070 KTDKENIEMTIENQBELRLGLDELKKQEI-----VAQEKHAIKKEGELSRCTD 1120
Db 1122 GKDSLGTVERTQDSSESEFK-----LEDPTVTPGLSFPNVPVCATPNSKILKE----- 1168
QY 1121 RLAEVTEKKEKESQ-----QLEKQOQLLVQEMSEMOKK 1156
Db 1169 -MCDVEQVL-SKTPPAPSPFDLPKLVHATEYQENKAPGKKK 1209

RESULT 11
US-09-592-054-2
; Sequence 2, Application US/09592054
; Patent No. 6440684
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Finer, Jeffrey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Wood, Kenneth
; TITLE OF INVENTION: No. 6440684el motor proteins and methods for
; FILE REFERENCE: 1016
; CURRENT APPLICATION NUMBER: US/09/592,054
; CURRENT FILING DATE: 2000-07-20
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1232
; TYPE: PRT
; ORGANISM: Human
; US-09-592-054-2

```

```

Query Match
Best Local Similarity 7.1%; Score 941.5; DB 4; Length 1232;
Matches 341; Conservative 227; Mismatches 480; Indels 255; Gaps 43;

QY 7 VAVCVVRPLNSREESLG-ETAQVYKTDNNVYQVDSKSFNDRFVHGHNETTKNVEE 65
Db 10 VRVALRCRPLVPKEISEGQCMCLSFVPGEQVVGTD--KSFTYDFVDFPSTQEVEFNT 67
QY 66 IAAPILDSAIQYNGTIFAYGQTASGKTYTMG-----SEDLGVIPRAIHDFOKIK 118
Db 68 AVAPLIKGVFGYNATVLAQGTSGKTYSMGAYTAQENETVGVIPRVIQILFKED 127
QY 119 KPDPREFLRVSMEIYNETITDLCCTQKMLPIREDVNRNVYVADLTVEEVYVYSEMA 178
Db 128 KKSDFEFTLKVSLEYNEYNEEILDLLCPREKAQINIREDPKEGKIIVGLTEKTVLVALDT 187
QY 179 LKWITKESRHYGETKMNORSRSHITPRMILESREKGEPSNCEGSKVSHNLVLDLAG 238
Db 188 VSCLEQGNRSRTVASTAMNSQSSRSHTIFTI-----SLEQGGKSKDNSSFR-SKLHLVDLAG 243

```

239 SERAAOTGAAGVRLKEGNCINRSLFILGOVKKISDQGVGFYINRDSKLTILQNSLGG 298
244 SERQKTKAEGDLKGINRGLCLGNVISALGDDKGGPAPYRDSKLTLLQDSLGG 303
299 NPKTRIICTIPV--SFDETLTALQFASHTAKYMNTPYVNEVSTDEALLKRYRKEIMDLK 356
304 NSHTLMIACVSPADSNLEETLTLRYADRAKIKNKPIVN-----IDPQTAEHLK 355
357 KQLEEVSL-----ETRAQAMEKDQALQLEEKD-----LLOKVONEK 393
356 QOVQOLVLLQAHGGTLPGSNABPSNLOSLMEKQ--SLVEENEKSLRCLSKAAGT 413
398 TRML-----VTSSSLTLOELKAKRRVTCWLGKINK-----MKNSTADQFNIPNITKT 450
414 AQMLERIINTEQANAKLEELRQHAACKLDLQKLVEITLEDQELKENVEIICNLQ 473
451 HKUSINLLREIDESCSDVFSNTLDTLSHIEWNPATK-----LLQEN 495
474 TQLS-----DETVAACMAAIDTAVEQAQVETSPETSRSSDAFTTQHALRQOMSKEL 526
496 IE-----SELNSLRADY--DNL-----VLDYEQRLTEKEMELKIK--EK 531
527 VELNKALAKELARKMTQDSLOPIQOYQODMKPELEVINLOKEKEELVLELQATK 586
532 NDLDEFEALERKTKQOQEMOLIHEISNLKLVKHEVYNQDLENELSKVLELREKEDI 591
587 KDAQAKLSERRRKLQLE--GQIADLKKL-----NEQSKLLKKESTERTV 633
592 KQLEYSIDSKLENIMDLSTYSIEDPKQMKQTLFDAETVAL---DAKESAFRLSE-- 647
634 SKLNQETRMKNQVQL--MRQMKEDAERFQWQKR--DKEVIQKERQKQYELLKLER 691
648 NLE-----LKEKMKELATTKQEMENDIOLYQSLQEAKKMQVDLEKELQSAFNEITKLT 702
692 NFQKQSNVLRKTEEAANKKDALQ-----KQREVADKEKETQSRMGCTA--- 740
703 LIDGKVPKLLCNLEKIGITDLQELKNEVEENALREEVILLSELK-SLPSEVERLRK 761
741 ---ARVKNWLGNEIEMVSTEBEAKRHLNLDLDRKILAQDVAQLKEKESGENPPKLR 797
762 BIQDSKELHIITSEKDLFSEVHKESRVQGLLEIGTKTDDATTQSNYKSTDEQFON 821
798 RTFSLTEVRGVQSESEDSITKQIESLETETEMFSAQIADLOQKLDDAESDR--PKQRWEN 856
822 FKTLMHDPQKYKMWLEENRNMQEIYNLSKAQKFDSSLGAKTSELYKTQELQEKTR-- 880
857 IATI---LEAKCAL-----KYLIGELVSKIQVSKLESLSKQSKTSCADMQKMLFEERNH 908
881 -----EVOERLNEMLQLEENRDSPLQTVREKTLITEKLOQTLREVKTLTQEKDDL 934
909 PAEIEETLQAEIVRMEQ-----QHPEKVLVLLSOLQSQOMAE----- 945
935 KQLESQIERQOLKSDIHDVTNMNIDTQOELRANALESLKQOETINTLKSLIS--EEVS 992
946 KQLESVSEKEQLOQS---TLKQDEELEKREVCQEQOQLRENEIITKQTLLOVAS 1001
993 RNLHMEENT-----GETKDBFQKQWGI--DKQDQLEAKNTQTLTADYK 1034
1002 RQKHLPKDTLLSPDSEFVYQPKPSRVKKEFLEQSDIEDLKCYSEHSVNEHEDGDGD 1061
1035 DNEIIEQQRKIFSLI-----OEKNE--LOQMLESVIAKEQ 1068
1062 DBEGDDEWKPTKLVNVRKNIQGCSCKGWGNKQCGCRKQKSDCGVDCDDCPTKCRNRQ 1121
1069 LKTDLKENTEMTIEQOEELRLGDELKQOEI-----VAQEKNHAIKKEGELSRIC 1119
1122 QGKDSLGTVERTQDSSESPK-----LEDPTFTVPLGSPFNPVCATNSKILKE----- 1169
1120 DRLAEVEEKLKEKSKQ-----QLEQEKQOQLLNVOEEMSEMQK 1156
1170 --MCDVEQVLSKTKTPAPSPFDLPKELKHVATEYQENKSGSKK 1210

RESULT 12
US-09-592-054-8
; Sequence 8, Application US/09592054
; Patent No. 6440684
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Finer, Jeffrey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Wood, Kenneth
; TITLE OF INVENTION: No. 6440684el motor proteins and methods for
; TITLE OF INVENTION: their use
; FILE REFERENCE: 1016
; CURRENT APPLICATION NUMBER: US/09/592,054
; CURRENT FILING DATE: 2000-07-20
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1234
; TYPE: PRT
; ORGANISM: Human
US-09-592-054-8

Query Match 6.9%; Score 924.5; DB 4; Length 1234;
Best Local Similarity 25.8%; Pred. No. 6.6e-36;
Matches 341; Conservative 228; Mismatches 398; Indels 357; Gaps 49;

QY 7 VACVVRPLNSRRESLG-ETAQVYWKTDNNVIYQVDSKSPNDRVPHGNETTKNVYEE 65
DB 10 VRVALRCRPLVPKEISEGCMCLSFVPGETQVVVGTD--KSTYDFVDFDPTCQEEVENK 67

QY 66 IAAPIIDSAIOGNGTIFAYGTAGTKTYTMG-----SEDHLGVIPRAIHDFOKIK 118
DB 68 AVAPLIGIFGYNATVLAAYQGTSGKTSYMGAYTAQENETVGIIPRIVQLLFKEID 127

QY 119 KFPDREFFLRVSYMEIYNETITDLCGQPKMKPLIREDVNRNVVADLTREVVVTSEMA 178
DB 128 QKSDFEFTLVSYLEIYNEEILDLCPREKAQINIRDPKEGIVGIVTEKTVLVALDT 187

QY 179 LKWTTKGSKSHYGETQANQSSSHTIFRMLESREKGEPSNCGSVKSHNLNVDLAG 238
DB 188 VSCLEQGNNSRTVASTAMNSQSSRSHAFIT---SLEQKKSDKNSSFR--SKHLVLDLAG 243

QY 239 SERAAOTGAAGVRLKEGNCINRSLFILGOVKKISDQGVGFYINRDSKLTILQNSLGG 298
DB 244 SERQKTKAEGDLKGINRGLCLGNVISALGDDKGGPAPYRDSKLTLLQDSLGG 303

QY 299 NPKTRIICTIPV--SFDETLTALQFASHTAKYMNTPYVNEVSTDEALLKRYRKEIMDLK 356
DB 304 NSHTLMIACVSPADSNLEETLTLRYADRAKIKNKPIVN-----IDPQTAEHLK 355

QY 357 KQLEEVSL-----ETRAQAMEKDQALQLEEKD-----LLOKVONEK 393
DB 356 QOVQOLVLLQAHGGTLPGSNABPSNLOSLMEKQ--SLVEENEKSLRCLSKAAGT 413

QY 394 IENLFRMLVTSSSLTLOELKAKRK--RRVTCWLGKINK-----MKNSTADQFNIPNIT 447
DB 414 AQMLERIINTEQ--VNEKLNALKEELRQHAACKLDLQKLVEITLEDQELKENVEIICNLQ 470

QY 448 TKTHKLSINLLREIDESCSDVFSNTLDTLSHIEWNPATKLLQENIESELNSLRADY 507
DB 471 QLTQLS-----DETVAACMAAIDTAVEQAQVETSPET-----SRSSDA 510

QY 508 DNLVIDYEQRLTEKEMELKKEKNDLDEFEALERK--TKKQOEMOLI--HEISNLKLVK 564
DB 511 FTTQHALHQAKSKEVEL---NNALAKLEALVRQOTQNDQLOQIPIQYQODNKNL-- 564

QY 565 HREVVNQDLENELSKVLELREKEDIKQLEIYDSQKLENIKMDLSYLSLESTEDPKQMK 624
DB 565 -----ELE-----VINQKEKEELVRELQ----- 583

QY 625 QTLFDAETVALDAKRESAFIRSENLEKMKELATTKQEMENDIOLYQSLQEAKKMQV 684
DB 584 -----TAKKNQAQAKLSEHR---RKLQ-- 603

```

QY 685 DLEKELASAFNEITKLTSLIDGKVPKDLLCNLEGGKITDLOKELNKE-----VENEALRE 741
Db 604 -----ELEGQIALDKKQKLNQSKLLKESTER 631
QY 742 EVILLS-ELKSLSEVERLERKEIQDKSEELHITSEKDKLFSEVHVHKSVOGLLEEIGK 800
Db 632 TVSKLNOEIRMMKQVRQVLMKQKDAEKFQWKQKDK---EVIQLKDRKQVLELLK 688
QY 801 TKDOLATQSNYKSTQDFONKTHLMDFFQKVKWVLEENRM-----NQEIVNLSKEA 854
Db 689 LE-----RNFQKSNVLRKKEEAAAANKRLKDALQKQREVADKRRKET 731
QY 855 QK--FDSLGALKTELSYKTOELQEKTRVQOERLNE-----EOLKEQLENRDS 901
Db 732 QSRGMEGTAARVRNWLGNEL-EVMVSTEEAKRHLNLDLEDRKILAQDVVLKKEKESREN 790
QY 902 PLOTVEREKLITIEKLOOTLEEVKTLTOEKDDLKQIQESQIQRDQOLKSDIHDVTANNID 961
Db 791 PPPKJRKCTFSLSEVHGQVLESDCIT-----KQI-ESLETEMELRSLAQIADLQOKLLD 843
QY 962 TOBOLNALESKHOHRTINT-----LKSKESEVSRNLHMEENTGETKDFQKQWVG 1014
Db 844 AESEDRP-----KQCHENTATILEAKCALKYLIGELVSSKIHV----- 881
QY 1015 IPKQDLEAKNTOTLTADVDKNEIIEQQRKIFSLIOEKNELOQMLSESVIAEKEQLKTDLK 1074
Db 882 --TKLENSLRQSKASCADM-----QKMLF--EEQNH-----SEIETELQ 917
QY 1075 -ENIEMTENOEBRLILGDLBELKQOQIIVAQEKQNHAIKKGELSRTCDRLAEEVKLEKS 1133
Db 918 ABLVRMEQOQKQVLYLVNLOQESQMAEKLQESASKEPQVST-----LQCODELEKX 973
QY 1134 QOLOEQOQOLLNVQOEMSEMOKKINENIENLKNELKNELTLE-----HMEERLELAOK 1187
Db 974 REVCEQOQOLL--QE-----NEI-----IKOKLILLOVASROKHLPNDTL--LS 1013
QY 1188 LNEYEEVKSITKERKVL--KELOKSPETERDHLRGYIRIEANGLTQKEELKIAHIH-LK 1245
Db 1014 PDSSPEYIPPKPSRVKEKFLQSDMI-----EDLKYSCHSVN 1053
QY 1246 EQHQB 1249
Db 1054 EHED 1057

```

RESULT 13

```

US-09-722-139-2
; Sequence 2, Application US/09722139
; Patent No. 6355471
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Freedman, Richard
; TITLE OF INVENTION: No. 6355471el motor proteins and methods for
; FILE OF INVENTION: their use
; FILE REFERENCE: 1055
; CURRENT APPLICATION NUMBER: US/09/722,139
; CURRENT FILING DATE: 2000-11-24
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: Human
US-09-722-139-2

```

```

Query Match 6.7%; Score 889; DB 4; Length 1375;
Best Local Similarity 26.2%; Pred. No. 3.4e-34;
Matches 380; Conservative 244; Mismatches 537; Indels 292; Gaps 56;

QY 6 AVACVVRVPLNSRESLGETAQVTKTDNNVIQV-----DGSKSFNDRVF 53
Db 3 SVKAVVRVPMNRNRREKDLKAKFIIQMEKSKTTITNLKIPGEGTGDGSGRETKTFTYDFSF 62

```

```

QY 54 HGNET-----TKNYVEEIAAIIIDSAIQYNGTIPAYQOTASGKTYTMGSEDHIGVI 106
Db 63 YSADTKSPDVVSOMVFKTLTGTDVWKSAPFEGYNACVAYQGTGSGKSYTMGNSGDSGLI 122
QY 107 PRAIHIDIFQKIK--KFPDREFLLRVSYMEIYNETITDLL-CGTOKMKPLIREDVNRNV 162
Db 123 PRICEGLFSINEITWDEASFTFVSYLIEYNERVDLLRRKSSKTFNLRVREHPKEGP 182
QY 163 YVADLTFEYVYTSEMALKWITKGEKSRHYGETKQNRSSRSHTIFRMILESREKGEPSNC 222
Db 183 YVEDLSKHLVQYGDVEELMDAGNINKTTAATGNDVSSRSHAFTTKFTQAKFDSMPCC 242
QY 223 EGSVKYSHLNLVDLAGSRAAQTGAAGVRLKECNCINRSLFILQGVIKKLSQGVGGFIN 282
Db 243 E---TVSKIHLVDLAGSERADATGATGVRLKEGNCINKSLVTLGAKKKQV-----FVP 292
QY 283 YRDSKLTIRIQLNSLGNPKTRIICTITP--VSDETLTALOPASTAKYMKNTPVVNEVST 340
Db 293 YRDSVLTWLLKDLGSGNSKTIATISPADVNYGETLSTLRYANRANKNIINKETINE-DA 351
QY 341 DEALLKRYKEIMDKKQLEEVSLQTRAKMEKQDLAQL-----LEEKDLLEKQVNE- 392
Db 352 NVKLIRELRAETARLKTLL-----AQNQIALLDSDPTALSMEEK--LQ--QNEA 396
QY 393 KIENLTREMLVTSSLT--LQOELKAKRKRRTWCGLK-----INK 430
Db 397 RVQELTEKWNKWNKTONILKEQTALRKEGTGVLDSELPHLGIDDDLLSTGIILYHL 456
QY 431 MKNSYADQFNPTNITTTKHLKISINLLREIDESVCSDESNTLDTLSEIEMNPATKL 490
Db 457 KEGQTVVGDDASTEQDIVLHGLDLE-----SEHCIFENIGTGTVTLPLSGQCS 506
QY 491 LNOENIESELNSLRADYDNLVDYEQLRTEKEMELKCKEKNDLDBEALERTKTKDQEM 550
Db 507 VNGVQI-VEATHLNOGAVILLGRTNMFNFHPKEAKLREK-----RKS 549
QY 551 QLIHEIS-NLKNLVKRE-----VYNQDLENELSKVELLREKEDQIKKLOVYDSOKL 603
Db 550 GLSSFSLSMTDLSKSRNLSAVMLYNGLEFEPQOREEL--EKLESKRKLEEME-EKQ 606
QY 604 ENIKMDLSYSLESIEDPKOMKOTLFDATVALDAKRESAFLRSENLEKKEKKELATIYK 663
Db 607 KSDKAELERMQOEV--TOKET-----EIVQIQIRQESLARRSPHINKLKLDAEKE 660
QY 664 QMENDIQLYQSOLAKKQWVD-----LEKEIQ-----SAFNEITKLTSLI 704
Db 661 KFEERLRQEQEIELQKQREBEETFLRQOELQRLKELNNKEAKKFOIFQELDQLOKEK 720
QY 705 DGKVPKDLLCNLEBKGITDLOKELNKEVEENEARLEEVILLSELKSLPSEVERLKEIQ 764
Db 721 DEQYAK-----LELE-----KGRLEQEQ--KEQVML-----VAHLEBQUR 753
QY 765 DKSEELHIITSEKDKLFSFV--VHKESR-VQGLLEIGTKDLDLATQSNYKSTQDFQCN 821
Db 754 EKQEMIQLLR-----GEVQVVEEKDLGEGIRSELVRKARA-----GGDEDGEE 800
QY 822 FKTLMDF-----EQYKVKVLEENRMNQEIIVNLSKEAQKFDSSLGALKTELSYKTOELQ 876
Db 801 LEKAQLRFFEFKRRQLVKLVNLEKDLVQOKDI-LKKEVQEQEEILECLKCEHD-KESRL 858
QY 877 EKTREVOERLNEBQLEQENRDSPLQTVREKTLITEKLOQOTLE-EVKTTITQKDDLK 935
Db 859 EKHDESVTDTVEVPQDFEFKIPVEYRLQYKERO-----LQYLLQNHLPTLLEBK---- 907
QY 936 QLOESLQI-ERDQLKSDIHDVTWNMIDTQEOLRNALBSLKQHOETINTL-KSKI SEVSR 993
Db 908 --QRAFELDRGFLSLD-----NTLYQVEKEWEEKBQLAQTQANANLQKLQATFEFTA 960
QY 994 NLHMEENTGETKD-----EFQKQKMGIDKKQDLKAKNTQTLTADVDKNEIIEQQRKIF 1046
Db 961 NIARQEEKVRKKEKILESRREKQKQREALERALARLERRHSALQRHSTLGTEIEQOKLA 1020

```

1047	QY	SLIQEKNELOQMLESVIAKSEQLKTKLDKGNIENTMIENQEBELRLUGDELKQOQIVAQEKX	1104
1021	DB	SLNSGSRQSGLOASLEAQEALQEKD-QERLEYEIQOLKQ-----KIYVDGVQKD	1070
1107	QY	HAIKQEGELSRCDRLAELVEEKLKESQOLQEQKQOQLLVQEESEMOKKINEIENLKNE	1166
1071	DB	H-----HGTL-----EGKVASSSLPVSAKSHLVP-----MDARINAY--IEEE	1108
1167	QY	LKNKELTLEHMETERLELAQKLNENYEEVKSITKERKVLKELOKSFETERDHLRGYIREI	1226
1109	DB	VQRRLODLHRVISEGCSTSDATMKDNEKLHNGTIQRKL-----KYELCRDLLCVLMPEP	1162
1227	QY	EATGLQTKBELKTAHILKHEQETIDELRSVSEKTAQIINTQDLEKSHTKLOEEIPTVLH	1286
1163	DB	DAACACANHPILQODLVQLSLDWKT--EIPDLVLPNGVQVSS-----KFQTTLVDMIVFLH	1215
1287	QY	EBEOLLPNVKVKSQETQETMNELELLTBSQSTTKDSTTLARIEMERLRLNEKFBQSEBEIKS	1346
1216	DB	GNMEV--NVPSLAEVQ-----LLYTTVKWGDSS-----GHDQCQS	1249
1347	QY	LTKERDNLKTIKE 1359	
1250	DB	LVLNTHIALVKE 1262	
RESULT 14			
US-09-721-832-2			
; Sequence 2, Application US/09721832			
; Patent No. 6399346			
; GENERAL INFORMATION:			
; APPLICANT: Beraud, Christophe			
; APPLICANT: Freedman, Richard			
; TITLE OF INVENTION: No. 6399346el motor proteins and methods for			
; TITLE OF INVENTION: their use			
; FILE REFERENCE: 1055			
; CURRENT APPLICATION NUMBER: US/09/721,832			
; CURRENT FILING DATE: 2000-11-24			
; NUMBER OF SEQ ID NOS: 4			
; SOFTWARE: FastSeq for Windows Version 4.0			
; SEQ ID NO 2			
; LENGTH: 1375			
; TYPE: PRT			
; ORGANISM: Human			
US-09-721-832-2			
Query Match 6.7%; Score 889; DB 4; Length 1375;			
Best Local Similarity 26.2%; Pred. No. 3.4e-34;			
Matches 380; Conservative 244; Mismatches 537; Indels 292; Gaps 56			
6	QY	AVAVCVRRPLNSREESIGETAQVYWKTDNNVIQV-----DGSKSFNFRVP 53	
3	DB	SVKVAVRVRPMNRREKOLEAKFIQMEKSKTTITNLKIPEGGTGDSGRERIKTFTDFSF 62	
54	QY	HGNET-----TKNVYEEIAAPIIDSAIQGVNGTIFAYGQTASGKTYTMMGSEDLGVY 106	
63	DB	YSADTKSPDYVSQEWVFKLTGVDVKSFAFEGYACVFAVGQTGSKSVYTMGNSGDSGLI 122	
107	QY	PRAIHDFQKTK--KPPDRFELRVSVMEIYNETITDLL-CGTQKMPLIIRDVNRNV 162	
123	DB	PRICEGLFSRINETTRDEASFRTEVSYLEIYNERVDRLLRRKSKTFLNRYREHPKEGP 182	
163	QY	YVADLTREVYVYTSMAKWIITKGEKSRHVGETKMNORSRSHTIFRMILESEKGEPSNC 222	
183	DB	YVEDLSKHLVQNYGDVEELMDAGNINRTAATGMDVSRSHAITFKTQAKFDEMP 242	
223	QY	BGSVKVSHNLVDLAGSERRAQTAGVRLKEGCNCNRSFLILGVYKKLSQGVGGFIN 282	
243	DB	E---TVSKHLVDLAGSERADATGATGVRLEGGNGINKSLVTLGAKKQV-----FVP 292	
293	QY	YRDSKLTILQNSLGCNPKTRICITLP--VSFDETLALQFASAKYMKNTYVNEVST 340	
293	DB	YRDSVLTWLLKDSLGCNKTMIATISPADVNYGETLSTLRVANRAKNNKPTNE-DA 351	

DB	1250 LVLLNTHIALVKE 1262	DB	607 KSDKAELERMQOEVB--TORKET-----EIVOLQIRKQOEESLKRSRFRHNIKIKDLAEKE 660
RESULT 15		QY	664 QMENDIOLYOSQLEAKKQMOVD-----LEKELQ-----SAFNEITKLTSLI 704
US-09-721-689-2		DB	661 KFEERLERQOEIELOKQROEETFLAVQOELOKELNNNEKAQFQIQELDQLOKQEK 720
; Sequence 2, Application US/09721689		QY	705 DGKVPKDLLCNLRLEGKITDLQKELNKEVEFEALREEVILLSELKSLPSEVERLKEIQ 764
; Patent No. 6440685		DB	721 DEQYAK-----LELE-----KKRLEEQE--KEQVNL-----VAHLEBQLR 753
; GENERAL INFORMATION:		QY	765 DKSEELHIITSEKDKLKFSEV--VHKESE--VOGLEEIGTKDLDLATTQSNYKSTDOQFQN 821
; APPLICANT: Beraud, Christophe		DB	754 EKQEMIQLLR-----GEVQWVEEKKDELEGRESLLRVKEARA-----GGDEGEE 800
; APPLICANT: Freedman, Richard		QY	822 FKTLHMDF-----EQKVMVLENERMNOQIIVNLKSEAQAQFSSLGALKTSLSYKTQELQ 876
; TITLE OF INVENTION: No. 6440685el motor proteins and methods for		DB	801 LEKAQLRFFEFKRRQLVKNLEKDLVQQKDI--LKKEVQOEQOEILECLKCEHD--KESRL 858
; TITLE OF INVENTION: their use		QY	877 EKTREVOERLNMEOLKEQLENRDSPLQTVVEREKTILITEKLOQTLF--EVKTLTQEKDDLLK 935
; FILE REFERENCE: 1055		DB	859 EKHDSEVTDVTEVPQDFEKIPVEYRQYKERQ-----LQYLLQNLHPLTLEEK--- 907
; CURRENT APPLICATION NUMBER: US/09/721,689		QY	936 QLQESLQI--ERDOLKSDIHDITVNMNITDQOIRNALLESKHOETINTLI--KSKISEVSR 993
; CURRENT FILING DATE: 2000-11-24		DB	908 --QRAFEILDRGFLSLD-----NTLYQVEKEMEKEQOLAQYQANANQLOKQATFEFTA 960
; NUMBER OF SEQ ID NOS: 4		QY	994 NLHMENTGETKD-----EFQKMWGIDDKODLEAKNTQTITADVKDNEIEQQRKIF 1046
; SOFTWARE: FastSeq for Windows Version 4.0		DB	961 NIARQEKVRKKEKEILESREKQREALRALARRRRHSALQRHSTGLTETEIEEQOKLA 1020
; SEQ ID NO 2		QY	1047 SLJQEKNEIQQMLESVIAEKEQLKTDLKENIEMTNIENQELRLLDGELKQOEIVAQEKN 1106
; LENGTH: 1375		DB	1021 SLNQSREGSLQASLEAEQEALEKD--QERLEYEIQQLAQ-----KIYEVDGVQKD 1070
; TYPE: PRT		QY	1107 HAIKKEGELSRTCDRILAEVEEKLKEKSQOLQEQKQOLLNVQEMSBSMQKINEIENLNKE 1166
; ORGANISM: Human		DB	1071 H---HGTL-----EGKVASSSLPVSAAKSHLVPL-----MDARINAY---IEE 1108
US-09-721-689-2		QY	1167 LKKNELTLEHMETERLELAQKLNENYEVKSIKERKVLKELQKSFETERDHLRGYIRI 1226
Query Match 6.74; Score 889; DB 4; Length 1375;		DB	1109 VORRLQDLHRVISEGSGSTSDATMKDKNLHNGTIQRKL-----KYELCRDLLCVLMPEP 1162
Best Local Similarity 26.24; Pred. No. 3.4e-34;		QY	1227 EATGLQTKEEELAHILHKEHOETIDELARSRSEKTAQIINTQDLEKSHTKLOEIPVLH 1286
Matches 380; Conservative 244; Mismatches 537; Indels 292; Gaps 56;		DB	1163 DAAACANHPLLQODLVQSLDWKT--EIPDLVLPNGVQVSS-----KFQTLVDMYFIH 1215
QY 6 AVAVCVVRPLNSREESIGETAQVYMKTDNNVIYQV-----DGSKSFNDRVF 53		QY	1287 BEQELLPNYKVSQETQETMNELELITQESTTKDSTTLARIEMERLRLNEKFOESSEELKS 1346
DB 3 SVKAVVRPMPNRRREKDLAKFILOMEKSKTIIINLKIPGGTGDSGRERTKTFYDFSF 62		DB	1216 GNMEV--NVPSLAEVQ-----LLLTYYTKVMGDS-----GHDQOCOS 1249
QY 54 HGNET-----TKNVYEEIAAPIIDSAIOQYNGTIFAYQTASGKTYTWMGSDHLGVI 106		QY	1347 LTKERDNLTKIKE 1359
DB 63 YSADTKSPDYVSQEMWPKTLGTDVVVSAFEGYNACVFAYGTQSGKSYTWMGSGDSGLI 122		DB	1250 LVLLNTHIALVKE 1262
QY 107 PRAHTDIFQKIK--KFPDRFLLRVSYMEIYNETITDILL--CGTQKMKPLIIRREDVNRNV 162			
DB 123 PRICEGLFSRINETFRWDEASFRTEVSYLETYNERVRDLRRKSSKTFNLVRHPEKGP 182			
QY 163 YVADLTBEVVYTSMAKWLKWTGKSKSRHYGETKMNQSSRSHTIFRMILSRKGPSPNC 222			
DB 183 YVEDLSKHLVQNYGDVEELMDAGNINRTAATGMNDVSSSHAIPTFKTFQAKFDSEMP 242			
QY 223 EGSVKVSHMLNVLLAGSERAQOAGAVRLKEGKNINRSFILGQVVIKSLSDQGVGFIN 282			
DB 243 E--TVSKIHLVDLAGSERADATGAVRLKEGKNINSLVTLGAKKQV-----FVP 292			
QY 283 YRDSKLTIRILQNSLGGNPKTRIICITIP--VSFDELTALQFASHTAKYMQNTPYVNEVST 340			
DB 293 YRDSVLTWLLKDSLGGNSKTIIMATIISPADVNTGETLSTLRVANRAKNIINKPTINE-DA 351			
QY 341 DEALKRYRKEIMDLKKQLEVSLETRAQAMKXDLQAL-----LESKDLLOKQVNE- 392			
DB 352 NVKLIREFRPAFIAPVTVI-----			

Search completed: July 29, 2004, 09:42:43
Job time : 77.178 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 09:28:21 ; Search time 39.7887 Seconds
(without alignments)
6437.961 Million cell updates/sec

Title: US-10-045-631B-88
Perfect score: 13329
Sequence: 1 MAEGAVAVCVVRPLNSRE.....SQPGPHASSGKDVPCKTQ 2663
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78: *
1: piri: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	13329	100.0	2663	1	S28261
2	3615.5	27.1	2954	1	T24156
3	1127.5	8.5	3259	1	A55539
4	1125.5	8.4	3225	2	I52300
5	1089	8.2	3187	2	JC5837
6	1056	7.9	1909	2	A45592
7	1048.5	7.9	1388	2	T30335
8	1026	7.7	2442	2	T08621
9	1004	7.5	823	2	T52425
10	980.5	7.4	1780	2	T17272
11	971.5	7.3	888	2	D56619
12	952.5	7.1	1231	2	A54803
13	946	7.1	1939	2	T18372
14	935.5	7.0	1226	2	I51617
15	929.5	7.0	1742	2	T49451
16	917.5	6.9	1225	2	A56514
17	900	6.8	2253	2	T30336
18	898	6.7	1459	2	T30196
19	896	6.7	1738	2	T14867
20	893.5	6.7	1957	2	T38077
21	891.5	6.7	1875	2	S38173
22	890	6.7	1979	2	C71622
23	887.5	6.6	1229	2	T48959
24	880.5	6.6	2829	2	A42771
25	878	6.6	2139	2	T18296
26	873.5	6.6	1827	2	T16270
27	867.5	6.5	963	1	A41919
28	867	6.5	1898	1	A45973
29	860	6.5	2331	2	T25410

ALIGNMENTS

RESULT 1

S28261
centromere protein E - human
N:Alternate names: centromere 312K protein; kinesin-related protein CENP-E
C:Species: Homo sapiens (man)
C>Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 19-Jan-2001
C:Accession: S28261
R:Yen, T.J.; Li, G.; Schaar, B.T.; Szilak, I.; Cleveland, D.W.
Nature 359, 536-539, 1992
A:Title: CENP-E is a putative kinetochore motor that accumulates just before mitosis.
A:Reference number: S28261; MUID:93024922; PMID:1406971
A:Accession: S28261
A:Molecule type: mRNA
A:Residues: 1-2663 <YEN>
A:Cross-references: EMBL:Z15005; NID:g29864; PIDN:CAA78727.1; PID:g29865
C:Genetics:
A:Gene: GDB: CENPE
A:Cross-references: GDB:361164; OMIM:117143
A:Map position: 4q24-q25
C:Superfamily: centromere protein E; kinesin motor domain homology
C:Keywords: ATP; coiled coil; microtubule binding; mitosis; nucleotide binding; P-loop
F:7-335/Domain: kinesin motor domain homology <KWOT>
F:86-93/Region: nucleotide-binding motif A (P-loop)
F:486-2183/Domain: coiled coil #status predicted <COI>
F:92/Binding site: ATP (Lys) #status predicted

Query Match	100.0%	Score 13329	DB 1	Length 2663
Best Local Similarity	100.0%	Pred. No. 0	Mismatches 2663	Conservative 0
Matches 2663	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1	MAEGAVAVCVVRPLNSRESLSGTAQYVKTNNVYQVDGSKSFNDRVFHGNETTK	60	
Db	1	MAEGAVAVCVVRPLNSRESLSGTAQYVKTNNVYQVDGSKSFNDRVFHGNETTK	60	
QY	61	NYVEIAAPIIDSAIOGNGTIFAYGQTASGKTYTMSGDHLGVIPIRAIHDFQIKKF	120	
Db	61	NYVEIAAPIIDSAIOGNGTIFAYGQTASGKTYTMSGDHLGVIPIRAIHDFQIKKF	120	
QY	121	PDREFLLRVSYMEIYNETITDLCGTQMKPLIREDVNRNVYVADLTVEEYVVTSEMALK	180	
Db	121	PDREFLLRVSYMEIYNETITDLCGTQMKPLIREDVNRNVYVADLTVEEYVVTSEMALK	180	
QY	181	WITKGEKSRHYGETNMQRSSRSHITFRMILESREKGEPSNCEGSKVSHLNLVDLAGSE	240	
Db	181	WITKGEKSRHYGETNMQRSSRSHITFRMILESREKGEPSNCEGSKVSHLNLVDLAGSE	240	
QY	241	RAAQTAAGVLEKGCNINRSLFILGVIKLSQGVGGFINRDSKLTIRLQNSLGNNP	300	
Db	241	RAAQTAAGVLEKGCNINRSLFILGVIKLSQGVGGFINRDSKLTIRLQNSLGNNP	300	
QY	301	KTRICTITPVSPFOTLTALQFASTAKYMKNTPYVNEVSTDEALLKRYRKEIMDLKKOLE	360	

301 KTRIICTITPVSFDETITALQFASATKYMNTPTVNEVSTDEALLKRYRKEIMDLKKQLE 360 Db
361 EVSLETPAQAWKQDLAQLEEDKLLQKVONEKIENTRMLVTSSSLTLOELKAKKRR 420 QY
361 EVSLETPAQAWKQDLAQLEEDKLLQKVONEKIENTRMLVTSSSLTLOELKAKKRR 420 Db
421 VTWCLGKINMKNYSYADQFNIPITNTTKTKLSINLLREIDESVCSSESDFVFSNTLDTLS 480 QY
421 VTWCLGKINMKNYSYADQFNIPITNTTKTKLSINLLREIDESVCSSESDFVFSNTLDTLS 480 Db
481 EIEWNPATKILNOENIIESELNSLRADYDNLVDYEQURTEKEEWEIKLKEKNDLDEFEAL 540 QY
481 EIEWNPATKILNOENIIESELNSLRADYDNLVDYEQURTEKEEWEIKLKEKNDLDEFEAL 540 Db
541 ERKTKDOEWMLIHEISNLKNLVKRVYVNDLENELSSKVELLREKEDQIKKLOEYIDS 600 QY
541 ERKTKDOEWMLIHEISNLKNLVKRVYVNDLENELSSKVELLREKEDQIKKLOEYIDS 600 Db
601 QKLENIKMDLSYSLESIEDPKMQKOTLFDABTVALDAKRESAFRSENLELKEKMKELAT 660 QY
601 QKLENIKMDLSYSLESIEDPKMQKOTLFDABTVALDAKRESAFRSENLELKEKMKELAT 660 Db
661 TYKQWENDIQYQOLAKKQWVDLEKELQSAFNEITKLTSLIDGKVPKDLLCNLELEG 720 QY
661 TYKQWENDIQYQOLAKKQWVDLEKELQSAFNEITKLTSLIDGKVPKDLLCNLELEG 720 Db
721 KITDQKELNEKEVENEAALREBVILLSELKSLPSEVERLRKEIQKSEBELHIITSEKDKL 780 QY
721 KITDQKELNEKEVENEAALREBVILLSELKSLPSEVERLRKEIQKSEBELHIITSEKDKL 780 Db
781 FSEVVHESRVQGLEEIGTKDKDLATTOQSNYKSTDQBFONFKTLHMDPEQKYMWLEEN 840 QY
781 FSEVVHESRVQGLEEIGTKDKDLATTOQSNYKSTDQBFONFKTLHMDPEQKYMWLEEN 840 Db
841 ERMQEIVNLSKBAQKFDSSIGALKTSLYKTOELQKTRVQBELNEMEOLEKOLENRD 900 QY
841 ERMQEIVNLSKBAQKFDSSIGALKTSLYKTOELQKTRVQBELNEMEOLEKOLENRD 900 Db
901 SPLQTVREKTLITEKLOOTLEEVKTLTQEKDDLKQLESQIQRDQKSDIHDVTNNMI 960 QY
901 SPLQTVREKTLITEKLOOTLEEVKTLTQEKDDLKQLESQIQRDQKSDIHDVTNNMI 960 Db
961 DTQEQLRNALSQKQOFTINTLSKISIEEVSRLNLMHEENTGETKDEFOQKQWVGIDKKQD 1020 QY
961 DTQEQLRNALSQKQOFTINTLSKISIEEVSRLNLMHEENTGETKDEFOQKQWVGIDKKQD 1020 Db
1021 LEAKNTQTLTADVKNONEIIEOQRKIFSLIOEKNELOQMLSVIAEKEOLKTDLKENIEMT 1080 QY
1021 LEAKNTQTLTADVKNONEIIEOQRKIFSLIOEKNELOQMLSVIAEKEOLKTDLKENIEMT 1080 Db
1081 TENQBELRLIGDELAKKQOEIVAOEKNHAIKKEGELSRITCDRLAEVEEKLKESQOLQEKQ 1140 QY
1081 TENQBELRLIGDELAKKQOEIVAOEKNHAIKKEGELSRITCDRLAEVEEKLKESQOLQEKQ 1140 Db
1141 QQLLVQEMSEMOKKINEIENLKNELKNKELTLEHMETERLELAOKLENTEYKSTIK 1200 QY
1141 QQLLVQEMSEMOKKINEIENLKNELKNKELTLEHMETERLELAOKLENTEYKSTIK 1200 Db
1201 ERKVLKELOKSFETERDHLRGVIREIATGLQTKELKTAHHLKHEOETIDELRSVSE 1260 QY
1201 ERKVLKELOKSFETERDHLRGVIREIATGLQTKELKTAHHLKHEOETIDELRSVSE 1260 Db
1261 KTAQIINTQDLKESHTKQEBIPVLHBEQELLPNVKVSETOETWNELELLTEQSTTKDS 1320 QY
1261 KTAQIINTQDLKESHTKQEBIPVLHBEQELLPNVKVSETOETWNELELLTEQSTTKDS 1320 Db
1321 TTLARIEMERLIRNKPQESQEBEIKSLTYERONLKTIKALEVKGHDOLKEHRETLAKIO 1380 QY
1321 TTLARIEMERLIRNKPQESQEBEIKSLTYERONLKTIKALEVKGHDOLKEHRETLAKIO 1380 Db
1381 ESQSKQESQINMKEKNDNETTKIVSMEQPKDQKDSALLRIEIMGLSKELQESHDBMKSV 1440 QY
1381 ESQSKQESQINMKEKNDNETTKIVSMEQPKDQKDSALLRIEIMGLSKELQESHDBMKSV 1440 Db

1441 AKKDDLOLQLOEVLOSSEDQLENIKIYAKHLETEELVAHCCLEKEQETINELRWL 1500 QY
1441 AKKDDLOLQLOEVLOSSEDQLENIKIYAKHLETEELVAHCCLEKEQETINELRWL 1500 Db
1501 SEKETETSTIQOLEAINDKLQNKIOEYIEKEBOLNIIKOISEQVENVNELKQFKEHRKAK 1560 QY
1501 SEKETETSTIQOLEAINDKLQNKIOEYIEKEBOLNIIKOISEQVENVNELKQFKEHRKAK 1560 Db
1561 DSALQSIESKMLEITNRLQESQEBEIQIMIKESEMKGVBQALQIERDQKLENTKEIIVAKM 1620 QY
1561 DSALQSIESKMLEITNRLQESQEBEIQIMIKESEMKGVBQALQIERDQKLENTKEIIVAKM 1620 Db
1621 KESQEKYQFLKMTAVNETQEKCEIEHLKEQEPQTKNLNLENTENIRLTQIILHENLEE 1680 QY
1621 KESQEKYQFLKMTAVNETQEKCEIEHLKEQEPQTKNLNLENTENIRLTQIILHENLEE 1680 Db
1681 MRSVTKERDDLSRVEETLKVERDQKLENRETITRDLEKOBELKIVHMHLEKHEOETIDKL 1740 QY
1681 MRSVTKERDDLSRVEETLKVERDQKLENRETITRDLEKOBELKIVHMHLEKHEOETIDKL 1740 Db
1741 RGVISEKTNEISNMQKOLEHSNDALKAQDLKIOBELRIAHMHLKEQOETIDKLRGVISEK 1800 QY
1741 RGVISEKTNEISNMQKOLEHSNDALKAQDLKIOBELRIAHMHLKEQOETIDKLRGVISEK 1800 Db
1801 TDKLSNMQKOLENSNAKLOEKIOELKANEHOLITLKKDVNETQKVSMEQOLKKQIKDOOS 1860 QY
1801 TDKLSNMQKOLENSNAKLOEKIOELKANEHOLITLKKDVNETQKVSMEQOLKKQIKDOOS 1860 Db
1861 LTLSKLEIENLNAQELHENLEEMKSVMKERDNLRRVEETLKLERDQKESLOETKARDL 1920 QY
1861 LTLSKLEIENLNAQELHENLEEMKSVMKERDNLRRVEETLKLERDQKESLOETKARDL 1920 Db
1921 EIQOELKTARMLSKHEKETVDKREKISEKTIQISDIQKOLDKDELOKKIOELQKKEKEL 1980 QY
1921 EIQOELKTARMLSKHEKETVDKREKISEKTIQISDIQKOLDKDELOKKIOELQKKEKEL 1980 Db
1981 QLLRVKEDVNMHSHKINEMEOLEKQEPNYLCKCEMDNFQTKKLHESLEIRIVAKERD 2040 QY
1981 QLLRVKEDVNMHSHKINEMEOLEKQEPNYLCKCEMDNFQTKKLHESLEIRIVAKERD 2040 Db
2041 ELRRIKESLKNRQOFTATLREMIARDQNHQVPEKELLSDDGQOHLMESLREKCSRIKE 2100 QY
2041 ELRRIKESLKNRQOFTATLREMIARDQNHQVPEKELLSDDGQOHLMESLREKCSRIKE 2100 Db
2101 LLKRYSEMDDHYECLNRLSLDLEKEIEFHRIMKKLYVLSVYTKIKEQHECINKFEMDF 2160 QY
2101 LLKRYSEMDDHYECLNRLSLDLEKEIEFHRIMKKLYVLSVYTKIKEQHECINKFEMDF 2160 Db
2161 IDEVEKQKELIKIHOLOQDCDVSRELRDLKLNQMDLHIFEILLKDFSESEFPKIKTEF 2220 QY
2161 IDEVEKQKELIKIHOLOQDCDVSRELRDLKLNQMDLHIFEILLKDFSESEFPKIKTEF 2220 Db
2221 QOVLNRRKEMTOFLEEWLNTFRDIEKLNKGIQKENDRICQVNNFNNRIIIMNNESTPFE 2280 QY
2221 QOVLNRRKEMTOFLEEWLNTFRDIEKLNKGIQKENDRICQVNNFNNRIIIMNNESTPFE 2280 Db
2281 ERSATISKWEQDLKSLKEKNEKLPKNYQTLKTSLSAQVNPPTQDNKNPHVTSRATOL 2340 QY
2281 ERSATISKWEQDLKSLKEKNEKLPKNYQTLKTSLSAQVNPPTQDNKNPHVTSRATOL 2340 Db
2341 TTEKIRELENSLHAEKESAMHKEKSKIEKQKELEVTNDIIAKQAKVHESNKCLKTKET 2400 QY
2341 TTEKIRELENSLHAEKESAMHKEKSKIEKQKELEVTNDIIAKQAKVHESNKCLKTKET 2400 Db
2401 IQVLQDKVALGAKPYKEEIEDLKMLVKIDLEKMKNAKEFEKEISATKATVYQKEVIRL 2460 QY
2401 IQVLQDKVALGAKPYKEEIEDLKMLVKIDLEKMKNAKEFEKEISATKATVYQKEVIRL 2460 Db
2461 LRENRRSQOQADTSVISEHTDQPSNKPLTCGGSGIVONTKALILKSEHRLKEIKS 2520 QY
2461 LRENRRSQOQADTSVISEHTDQPSNKPLTCGGSGIVONTKALILKSEHRLKEIKS 2520 Db

Db 1561 IDTTLKHSDTAQLOKTOQLOLAKNLAIAASDNCSPITOKE-----TSADCVHPLEKIL 1617
QY 1425 GLSKRLOESDEKMSVAKKDDLQRLQEVLOSQDQ-----KENIKIIVAKHLETEE 1477
Db 1618 LLEELHOKTNEQKLEHKNELFOAQUELKEVEHLMKSMISKSLSLESHKEKHDTQ 1677
QY 1478 ELKVAHCLKEQEETINELRNLSKE-----TEISTIQKOLEAINDKLNKIQBIYEKE 1533
Db 1678 QL-----LALAQOMQVVQEKELQQTTHLTAEDVHLKENIE-LGLNFKNEAQOKTKEQ 1732
QY 1534 QL-----NIKOISQEVQNVN-ELKQPEHRAKADALQSI-----ESKMLELTNRLQESQEIQI 1587
Db 1733 CLNENELQSQHRLQCEIEMKSLDKESALETLKESQKVINL-----NOEMEM 1785
QY 1588 MIKEKEMKRVQOALQIETDQLKENTKEIVAKMESQE--KEYQFLAKMTAVNETQEKWCE 1645
Db 1786 VMLEBESLKSQRTVIARDDQLQDQLRESVMSIETQDRLKAKAEALQOQKDKVQLTSQ 1845
QY 1646 IEHLKEQFETQKLNLENIETNIRLTQILHENLEEMRSVTKERDDLSRVEETLKVERDQL 1705
Db 1846 ISVLQEKISL-----LEN-----QWLY-NVATVKETLSERDQLNQSQKHLSEIETL 1891
QY 1706 KENIRETITRDLEKQBELKIVHMLKHEQFTID--KLRGIVSEKTNIEISNMOKDLHNS 1762
Db 1892 SLSLKEK-EFALQAEKDK-----ADAARKTIDITEKISNTEEBQLLOQATNKLKTL-YER 1944
QY 1763 DALKAQDLKIQBELRIAHMLKEQOFTIDKLGIIVSEKTDKLSNMOKDLENSNAKLOEKI 1822
Db 1945 ESL-----IQCKEQLALNTEHLRETLKSKDLALGWQEERDEAANKVIALTEKMSLEEQI 2000
QY 1823 QE-----LKANPHQIITLKKQVNETQKKVSEMEQKKQIKQOQSLTISKU-----BIE 1869
Db 2001 NENVTLKEGEGEXETPYLRPSKQSSQSSQMEELRESLTKDQLQLEAEKISEATNEIK 2060
QY 1870 NL-----NLAQELHENLEEMSVMKERNIARVEETLKLRDQKESIQETKARDLEIQ 1924
Db 2061 NLTKISLSLEIILQNASILNEAVSERENLRHSQQLVSELEQLSLTL-----KSRDHAFAP 2117
QY 1925 ELKTARMLSKHKEVTVDKLRKISBKTTQISDIQKDLK--SKDELQKTKIQELQ----- 1976
Db 2118 -----SKREK-----DEAVNKIASLAEIKLITKEMDEFRDSKESLQESQSHLSEELCT 2166
QY 1977 -KKELQLLR-VKEDVYMS-HKKINEMQIKQFEPNYLCKCEMDNFQUTKKLHESLEIR 2033
Db 2167 YKTELQMLKQKEDINNKLAEKVEDEL-----LOHLSLSLEQLDQIQ 2210
QY 2034 IVAKERDELRLKESLKE-ME-RQFIATREMIARDRQNHQVKEPKRLISDGOQHLMBSLR 2092
Db 2211 MELR-NEKLRNYELCEKMDIMEKEISVLRLM-----ONEPOOE--EDDVAERMDILE 2259
QY 2093 EKCSRIKELLKRYSEM-DDHYECLNRLSLDLKEIEFHR----- 2130
Db 2260 SRNQEIQELMEKISAVYSQHTLLSSLSSELQKETEAKHCHMLNIKESLSLTSRSPGSL 2319
QY 2131 -----IMKKLYVLSYVTKIEBOHECINKFEMFIDEVEKQKELLIKIQL 2177
Db 2320 QTEHVKLTQLQTLANKPKVW--YRTAAVKEDHSLIKOYEKDLAAEQKHDELRLQLOCL 2377
QY 2178 QDQ-----CDVPRELRDLKLNQMDL-----HIBEILKDFESE--FPSIKTEFOQVLS 2225
Db 2378 EQHGRKWSDSASEELFCBIEFLNELLFKKANIQSVQDDFSEVQVFLNQVGTLOFELE 2437
QY 2226 NRKEMTQPLEEWLNTFRDIEKLKNGIQKENDRICQVNNFNFRITAINNESTEFERSAT 2285
Db 2438 HKKGFQWLEEFGLHVDKAKLSEGQENRRIASTIQLTKRLKAVVQSKI---QREIT 2494
QY 2286 I-SKEWQDLKSLKEKNEKLFKNYQTLTKSLASGAQVN-----PTTQDNKNPHVTSRA 2337
Db 2495 VVLNQPEAKLOEKKEQKELMRMRHEHGPSSASVMEENARLLGILKTQDE-----SKK 2548
QY 2338 TQLTTEKIRELNSLHAKESAMHESKTIKMKQKELEVNTDIAXU-----QAKVHESNKCL 2394
Db 2549 LQ-----SRIKMLENELNDVDDAMHKEKVAILODKL-LSRNAEAEALNANQVKLTKKQDNL 2604

QY 2395 EKTERTIOVLQDKVALGAKPYKEEIEDLKMVKIDLEKMKNAKEFEKEISATKATVEYO 2454
Db 2605 QAAMKEIENLQKVVAGAVPYKEEIDNLKTKVKIEMEKI KYSKATQEIYAILKSCLEDK 2664
QY 2455 KEVIRLLRENLRSSQQAQDTSVISEHTDPOPSNKPCLTCGGSGGIVQNTKALILKSEHRL 2514
Db 2665 EGBLRRUKEELRAQADNDITVCVPKDYQKASFPVTCGGSGGIVQSTAMVLQSEKAAL 2724
QY 2515 EKESIKLKQONEQL-----TKQKNELLSNNQHL-SNEVKTWKERTLKREAHQV 2562
Db 2725 ERELSHYKVKVYHLSRTMSSSEDRKTKAKSDAHSSHTGSHRSGPHKTETIR---HGPV 2781
QY 2563 TCE-----NSPK-----SPKVTGT-----ASK 2579
Db 2782 TPERSEMPSLHLSGPKKSESSTKRVVSPNRSEIYSQVLMVPGKTMHKLHILSPKVLGHK 2841
QY 2580 KKOITP-----SCKERNLQDPVPKESPKSCFF 2607
Db 2842 KRALSPNRSEMPHTOHVISPCKTGLHKNLTBSTLFDNLSSPCQKQVQENL--NSPKGKLF 2899
QY 2608 DSRKSLSPSPHPVRYFNDNSISGLCPFQVONAGAEVDSQPGPWHASSGKDVPECKT 2662
Db 2900 DVKSKMPY-CPSQPFQNSKLGDFSELNTAESNDKSOAENWVWEAKKETAPECKT 2953

RESULT 3

A56539

Giantin - human

N:Alternate names: macrogolgin

C:Species: Homo sapiens (man)

C>Date: 19-Oct-1995 #sequence_revision 26-Jan-1996 #text_change 10-Dec-1999

C:Accession: A56539; S37536

C:Seelig, H.P.; Schranz, P.; Schroeter, H.; Wiemann, C.; Griffiths, G.; Renz, M. Mol. Cell. Biol. 14, 2564-2576, 1994

A:Title: Molecular genetic analyses of a 376-kilodalton Golgi complex membrane protein A:Reference number: A56539; MUID:94187728; PMID:7511208

A:Accession: A56539

A:Molecule type: mRNA

A:Residues: 1-3259 <SEE>

A:Cross-references: EMBL:X75304; NID:9405714; PIDN:CAAS3052.1; PID:9405715

C:Genetics:

A:Gene: GDB:GOLGB1; GCP: GCP371

A:Cross-references: GDB:454958

A:Map position: 3q13.31-3q13.31

C:Superfamily: giantin

C:Keywords: coiled coil; Golgi apparatus; transmembrane protein

P:3238-3254/Domain: transmembrane #status predicted <TMN>

Query Match 8.5%; Score 1127.5; DB 1; Length 3259;

Beat Local Similarity 20.9%; Pred. No. 1.2e-22;

Matches 643; Conservative 596; Mismatches 1011; Indels 831; Gaps 143;

QY 87 QTASGKTYTMGSEDLHGVIPRAIHDFQIKKFPDREFLRVSYMEIYNITDLCOT 146

Db 447 ETASQTSPPDVNVEGTQVTEENIASLQKRVVELENEKALLSSIE-----LEELKAEN 501

QY 147 QOKKPLIITREDVNRNVYVADLVEEYVYVYSEMAKMITKGEKSRHYGEIKMQRSSRH-T 205

Db 502 EKLSQITLLEAQNRTEADREVSBISIVDIANKSSSAEES---GQDVLENTFSQKHKE 558

QY 206 IFRMILESEKCEPNSCEGSKVSHNL-----VDLAGSERAATGAAGVR 251

Db 559 LSVLLEMKBAQE-----EIAFLKQLQKRAEADHEVLDQKEMQMEGEGLAPIK 610

QY 252 LKEGCNINRSLFLGQVINKLSD-GQVGGFINRYSKLTIRLQNSLGGNPKTRIIITPIP 310

Db 611 MK-----VF-----LEDTGDFPLMPEEESLPAVEKQASQTEHQSR---TSEE 651

QY 311 VSFDETLTALQFASTA---KMKNTPTVNEVSTDEALLKRYKKEIMDLK-----KQLE 360

Db 652 ISLND--AGVELKSTNQDGDKSLSAVFDIGQCHQDE--LERUKSQIILELNFHKAQEY 707

Db 2649 SSSQKRIAELEBEELVVCVQKAAKVGIE--DKLKKELHLLHHDAGIMRNTEETAERVA 2706
Qy 2188 --LRDL--KLQNDLHIEE-----LKDFSEFFPSIKTFQOVL 2225
Db 2707 ELARDLVEMEQKLLMVTKENKGLTAQOSFRSMSSLQNSRDHANEELDELKRYDASL- 2765
Qy 2226 NRKEMTQFLBEWLNTRFDIEKLKNGIQKENDRICOVNFFNNRIIAIMNDESTEPERSAT 2285
Db 2766 --KELAQKQGL--LNKRDALLSETAFSMN-----ST--EENSLS 2801
Qy 2286 ISKEWEOQLS-----LKEKNEKLPKNYQTLKTSLASGQVNPFTQDNKNPHVTSRAT 2338
Db 2802 HLEKLNQQLSKQKQHLHLSQLSDESYNQVQSFKAMAS-----LONERDHLWNE-- 2851
Qy 2339 QLTTKEIRELE-----NSLHEAKESAMHESKIIK-----MOKELEVT 2376
Db 2852 ---LEKFRKEEGKQKRSNAQSPSTPAEVOSLKKAMSSLQNSRDRLKELKLNQOQYLQIN 2908
Qy 2377 NDI--IAKQAKVHESKNCLEKTKETIQVLQKVALGAKPYKEIEDLQKLVKIDLEKM 2434
Db 2909 QETELHPLKAQLOEQ--DKTK-AFOIMOELRQENLSWQHHLQRLMEKSSWEIHER 2964
Qy 2435 KNAKEPEKEISATKATVEYQKEVIRLARENLRQQAQDTSV-ISEHTDPOPSNKELTCG 2493
Db 2465 RMKEOYLMAISDKQOQLSHLQNLRELSS---SOTQPLKVQYQKASPETISASP---D 3018
Qy 2494 GSGIVQNTKAL-----ILKSEHRLKEKESKLQKQNEOLIKQKN-----ELL 2536
Db 3019 GSONLVYETELRTQLNDSLKEIH-QKELRTQQLNSNPSQLLEBKNTLSIQLCDTSQSLR 3077
Qy 2537 SNNQH---LSNEVTKWERTKRAHKQVTCENSPKPKVTGTASKKKQIIPSCQKERNL 2593
Db 3078 ENQCHYGDLNHCRAVLEKQVQLOA-GPLNIDVAFGAPQKNGVHRK----- 3123
Qy 2594 QDPVPKSPKCFDPSKSLPSHPVRYFNFSSILGLCEV--QWAGAESVDS----- 2644
Db 3124 SDPELRFPQSFSEAOQOLNTRQEV---NELRKLLEERDQKVAENALSVABEQIR 3179
Qy 2645 --QPGMHPASSGKDVPECKTQ 2663
Db 3180 RLEHSEWDSSTPIIGSGTG 3200
RESULT 4
152300
N:Alternate names: gcp372
C:Species: Homo sapiens (man)
C>Date: 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 20-Aug-1999
C:Accession: I52300
R:Shoda, M.; Misumi, Y.; Fujiwara, T.; Nishioaka, M.; Ikehara, Y.
Biochem. Biophys. Res. Commun. 205, 1399-1408, 1994
A:Title: Molecular cloning and sequence analysis of a human 372-kDa protein localized in
A:Reference number: I52300; MUID:95100974; PMID:7802676
A:Accession: I52300
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-3225 <RES>
A:Cross-references: GB:D25542; NID:9662389; PIDN:BAA05025.1; PID:g808869
C:Superfamily: giantin
Query Match 8.4%; Score 1125.5; DB 2; Length 3225;
Best local Similarity 21.3%; Pred. No. 1.3e-22;
Matches 599; Conservative 529; Mismatches 963; Indels 718; Gaps 119;
Qy 251 RLKEGCNINRSFILGVQIKSLSDQGVGFI---NVRDSKLTPI-----L 292
Db 680 KAKEISLNQ-----LIEEFKNADNNSAFTALSERDQLLSQVKLSWVTELRAQVKQL 735
Qy 293 QNSLGCNPKTRICITPVSFDETLT---ALQFASATKMYNTPVNEV-----STD 341
Db 736 ENWLAERQRRLDYESQTAHNLNLTQIHLISLEAKSKVKIEVLQNELDQVQLQFSEQ 795

Qy 342 EALLKRYRKEIMDLKKQLEEVSLTRAQAMEKQDQALQLEEKDQLQKQVONEKIEINLTML 401
Db 796 STLIRSQSQONKQSEVLEGAERVRISSKVEELSQALSQKEL-----EITMDQLL 848
Qy 402 VTSSS--LTQOELKAKRKERVWCLGKINKMKNSTADQFNIPNTIITTKHLKLSINLLR 459
Db 849 LEKGRDVTETQOITEEKQOQVTEISFSMTKQWQLN-BEKFSLGVEI--KTLKEQLNLLS 905
Qy 460 EIDES-----VCSSESDVFSNTLDTLSEIEMNPATKLNQENIESEINSIRADYD----- 508
Db 906 RABEAKKEQVEEDNEVSSGLKQNYDEM--SPAGQ-ISKELQHEFDLLKKENEQKRRKLQ 962
Qy 509 -----NLVLDYEQLEATE-KEEMELKKEKNLDLDEFALEK-----TKK 546
Db 963 AALINRKELIQRVSRLEBEELANLKDESKKEIPLSETERGEVEEDK--ENKESYSEKCVTSK 1020
Qy 547 DQEMOLITHEISNLKLVKREVVNQDLENELSSXV-----ELLREKEDQIK 592
Db 1021 QOEIY-----LQOTISEKEVELOHIRKOLEEKLAAEQFOALVKQMNQTLQDKNQID 1075
Qy 593 KLQBYIDS-----QKLENIKMDLSYLESIEDPKMQKTLTDAETVAL----- 635
Db 1076 LLQAEISENQAIQKLTITSNTDAS-----DGDSVALVKETVWISPPCT 1118
Qy 636 -----DAKRESAFRSNLELKEKMKELATTVKOMENDIQLVQSOLEAKKQMOVDLE 687
Db 1119 GSSBHWKPELEETILALEKEKQLOKLOEALTSRKAILKKAQEKERHLRELKQKQDDY 1178
Qy 688 KEQSAFNEITK-----LTSIDGKVP----- 709
Db 1179 NQLQEQDEQSKENENIGDQLRQLQIVRESIDGKLPSDQOQSCSSPTGLEELPKATE 1238
Qy 710 -----KDLCC-----NLELEKIDTLOKEL 729
Db 1239 QHHTQPVLESNCLPWFPSHSEDASALQGTSAQIKAKLKEIAEKVELELKVSTTSSEL 1298
Qy 730 NKEVEENAELEEV-----ILLSELKSLPSEVERLKEITQDKSE-----ELHI 772
Db 1299 TKSEEVFQLOEQINKQGLEIESLKTWSHEAEVHAEISLQKLESSQLQIAGLEHLELOP 1358
Qy 773 ITSEKDKLFSVWHKESRVQGLLEEIGTKTDLLATTSQNSYKSTQDQFQNFKTLHMDFPQK 832
Db 1359 KLDELQKLSKKEEDSVLSQQLSE-----KEAALTQIQTIEEIEQ-DLIKALHTQLEMQ 1412
Qy 833 YKMVLEENERNQOIVNLSKEAQKFDSSLGALKTYSYKTQELQEKTEV-----QERLNE 888
Db 1413 AK--EHDRIKQLQVELCEMKQK-PEEIG-----EESPAKQIQKLOQALISRKEALKE 1464
Qy 889 MEQLKEQLENRDSPLQTVREKKTILITEKLOQTLEBVKTLTQKDDLKQQLQESLQJERDQ 948
Db 1465 NKSQEEELSARG-----TIER-----LTKSLADVESQVSQNKKEKTVLGRFALLQERDKL 1517
Qy 949 KSDIHDVTVMNIDTQEOURNALIESIKQHOETINTLKSKEISFVS--RNLHMEENTGETKD 1006
Db 1518 ITEMDSILEN-----QSLSSSCSELKALEGLTEDKEKLVAIESLKSKEIAEST----- 1568
Qy 1007 EFQKQWGDGDKQDLKAKNTQTLADVKDNEIIEOORKI FSLIQKNEKLEQMLQESVIAEK 1066
Db 1569 EWOEKHKELOKYEILLQSYENVS-----NEAERLQHVVEAVROEQKELYGKLRSTEANK 1623
Qy 1067 EOLKTDLKNTEMTIENOEELRLGLDELKQOEIV-AQEKNAIKKE-----GELSRTC-- 1119
Db 1624 KETEQLOQAEQOEMEEMKEKMKFPK--SKQKILELEBEENDRLRAEVHPAGDTAKECME 1681
Qy 1120 -----DRLAEVEEKLKEKSQLQKQOQLNVOQEMSEMOKKINIEINLKNELK--NK 1170
Db 1682 TLLSSNASMKEBELERVKMEYETLSKKFQSLNSEXKDSLSE-----EVODLKHQIEGNVSK 1735
Qy 1171 ELTLEHMETERLELAQKLNENYEEVKSITKERVKLKELQKSPETERDHLRGYIREATG 1230
Db 1736 QANLE--ATEKHD--NOTNVTGEGFQSPGTE-----BODSLSMSTR---PTC 1777
Qy 1231 LQTKELKIAHHLKHEQBTIDELRRSVSEKTAQIINTQDLEKSKTLOEIPVLHEEOE 1290

Dd	1778	SESVPASANPANSKDFSSHDEINNYL-----QQID-----QKERRIAGLEEKQ	1823
Qy	1291	LLPNVKYSVETQETMNELELLTEQSTTKDSTTLARIEMERLINERKQFQESQESIKUTKE	1350
Dd	1824	---KNKEFSQTLN-NEKNVTLISQISTKDG-ELKWLQEEVTKMNLNQIQOELSRVTK-	1876
Qy	1351	RDNLKTIKEALEVVKHDQIKHEIRITLAKIQESQSQKQSQSLNMKKNQNTTKVISEMEQFK	1410
Dd	1877	-----LKETAEEKDDLEERLMMQMLAELNGISGNYCQDVTDQAIKNEL-----	1919
Qy	1411	PKDSALLRIETEMGLSRLQESHDENKMSVAKEDKDLQRLQEVLOSSEDQIKENIKEIVA	1470
Dd	1920	-----LESEMKNLKCYSLEEEKQOLVKEKTV-----ESEIRKBYLEKIQG	1962
Qy	1471	KHLE-TEEEUKVAHCCLKQBEETINELRVNLSEKETEISTIQKLEAINDKLQNKIQEI	1528
Dd	1963	AQEPFGNKSNAKELQELLEKQKQVQKQKDXQDCIRYQEKISALERTVKAL-----EF	2013
Qy	1529	YEKERQNLNIQISVEQVNNVELKQFKERRAKUSALQSIE-----SKMLEL	1574
Dd	2014	VQTESQ---KOLEITKEN---LAQVHRKKAQAEASFPKVLDDTQSEARVLADNLKL	2067
Qy	1575	TNRLQESQEEIQIMKEKEE--MKRVOEALQIERDOLKENTKEIVAKMSESKEKYQFLK	1632
Dd	2068	KKELQSNKESVKSQMKQKDEDLERLSQA--BEKHKE-----KKNQMEKLDALR	2115
Qy	1633	MTAVNETQKMCETEHL--KEQFTQKL--NLNFIETNIRLTQIHLNLEENRSVTKE-	1687
Dd	2116	REKVH-LEETIGEIQVTINKDKVEQOQLENLDSTVTQLAAFTKSMSSLDQDDRVIDEA	2174
Qy	1688	-----RDDLRSVEETLKVBERQ--LKENLAEITTRDLEKQEBELKIVIMHLKEH----	1733
Dd	2175	KKWERKSDALQKBEIRILKNCVSLKQDLQMSIH-----MEEIKI-NISELHDKOI	2229
Qy	1734	-----QETIDKLRGIVSEKTEINSMOKLLEHSNDALKAQDLKIQELRIAHM	1781
Dd	2230	WESKAQTEVQLQKQVCDTLQ-----ENKELLSQLEFTRHLHYSSQNELAKJISELSKLD	2285
Qy	1782	HLKEQOETIK-----LKGIVSEKTKDLN-----MOKDLENSNAKLQEKIOELKA	1827
Dd	2286	QLTDLSNLSBKCKEKGKNGLEGIIRQEAQIONSKFYSVEQLETDLOASRELTSLRLEEINM	2345
Qy	1828	NEHOLITLKQDVNET-----QKKVSEWOLKKQIKQOOL-----	1861
Dd	2346	KEQKISLLSGKEEAQVATAELRQOHDKEIKLENLLSQEEENIVLBEENKQAVDTN	2405
Qy	1862	-----TLKLEIENLNAQELHENLEEMKSVYMERDNL-----RRVEE--TLKLERDOLKE	1910
Dd	2406	QLMETTLTKIKENIQOKAQLDSPVKSMSSILONDRDRIVGDYQOLEERHLSIILEKQDL--	2463
Qy	1911	SLOETKARDLEIQOELKTARMLSKHEHETVDKUREKISEKTIQISDIQKDLKSKDELQK	1970
Dd	2464	-IOEAAENAKNLKBEIRGLR-----SHMDILNSENAKLDAELIQYREDLNQVITIKDSQOK	2518
Qy	1971	KIQEOLKQLQLLRVKEDVNMHSKKINEMOLKKQFEPNYLCKCENDMPQLTKLHESLE	2030
Dd	2519	QLLEVO-----LQONKELENKYAKLEKKEUSEAN-----EDLRRSFN	2557
Qy	2031	EIRIVAKERDELRIKESLKMVERDQFIATLRMIARDRONHGVKPEKRLLSDGQOHLMES	2090
Dd	2558	ALQ---EEKQDLSKETESLKVSISQL-----TRQVTAQOEGTL---GLYHAQIK	2601
Qy	2091	LR-EKCSRIKELL---KRYSEMDDHYECLNRLSLDLEKIEIFHRIMKKLY-----VL	2139
Dd	2602	VKEEVRHRLSALFSSQKRIAELEEBELVQCEAAKKGVEIE-DKLKXELKHLHHDAGIM	2660
Qy	2140	SVYTKIKEQECINKEFMDFIDEVEKQKELLIKIQLHOQDCDVPSS--RELADKLNQNM	2197
Dd	2661	RNETETAER---VLAELADLVB--MEQKLLMVTVENKGLTAQIQSFGRSSSL---QNS	2712
Qy	2198	DLHIEBILKDFGESEFPSTKTEFOQVLSNRKEMTQFLEEWLNTRFDIEKLXNGIOKENDR	2257

Db 2713 -----RDHANEIDELKKYDASL-----KELAQLKEQGL-----LNRRDA 2749
 Qy 2258 ICQVNNFFNRIATMNSTEERSATISKEWODLK-----LKEKEKLFPKVQT 2310
 Db 2750 LLESETAFSWN-----ST--EENSLSHLKINQQLLSKDEQLHLSSOLEDSYNQVS 2799
 Qy 2311 LKTSLASGAQVNPTTQDNKNPHVTSRATQLTETKIRELE-----NSLH 2353
 Db 2800 FSKAWAS-----LQNERDLWNE-----LEFRKSEEGKORSAAQPSTSPAEVQSILK 2846
 Qy 2354 EAKESAMKESI-K-----MKOLEVTNDI--IAKLQAKVHENSKCLEKTETIQVLQD 2406
 Db 2847 KAMSQNDRLRDLKELKNLOOQLNQINEITELHPKLQAQEYQ---DKTK-AFQIMOE 2902
 Qy 2407 KVALGAKPYKEIEDLKMKLVKIDLEKMNAKEFEKSATKATVEVQKEVIRILRENLR 2466
 Db 2903 ELRQENISWQHLEOLRMEKSWIEHRMKQBQYMAISDKOOQLSHONLIRELRS-- 2960
 Qy 2467 RSQAQAQTSV-ISEHTDPQPSNKELTCGGSGGIQNTKAL-----HLKSEHILEKEIS 2519
 Db 2961 -SSQTPLKVQYORQASPETSASP---DGSQLVYETELLARTQLNDSLKEIH-QKELRIQ 3015
 Qy 2520 KLKQONQLIKQN-----ELLSNNQH---LSNEVKTWKERTLKREAHKWTCSE 2565
 Db 3016 QLNSNFQSLLEEKNLTSLQCDTSQSLRENQQHYGDLLNHCAVLEKQVBLQA-GPLNID 3074
 Qy 2566 NSPKSPKVTGTASKKKQITPQCKERNLPVPKESPCKSCFEDSRSKSLSPHPVRFDN 2625
 Db 3075 VAPGAPQEKNGVHK-----SDPEELREFQSFSAQQLCNTRQEV----N 3117
 Qy 2626 SSLGLCPREV--QNAAGSVDS-----QPGFWHASSGKDVPCEKTTQ 2663
 Db 3118 ELRKILLEERDQRVAENALSVAEEQIRLRLEHSEWDSSRTPIIGSCGTQ 3166

RESULT 5
 JC5837
 C:364K Golgi complex-associated protein - rat
 C:Species: Rattus norvegicus (Norway rat)
 C>Date: 05-Mar-1998 #sequence_revision 13-Mar-1998 #text_change 20-Jun-2000
 C:Accession: JC5837
 R:Toki, C.; Fujiwara, T.; Sohda, M.; Hong, H.S.; Misumi, Y.; Ikehara, Y.
 Cell Struct. Funct. 22, 565-577, 1997
 A>Title: Identification and characterization of rat 364-kDa Golgi-associated protein rec
 A:Reference number: JC5837; PMID:98093490; PMID:9431462
 A:Accession: JC5837
 A>Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-3187 <OK>
 A:Cross-references: DDBJ:D25543; NID:G516825; PIDN:BAA05026.1; PID:G516826
 C:Comment: This protein plays a role in the formation and maintenance of the characteris
 C:Superfamily: Giantin
 F:49-549,624-1176,1238-1707,1763-3114/Domain: coiled-coil leucine zipper #status predict
 F:3165-3187/Domain: membrane anchor #status predicted <MAD>

Query Match 8.2%; Score 1089; DB 2; Length 3187;
 Best Local Similarity 20.9%; Pred. No. 1.2e-21;
 Matches 634; Conservative 558; Mismatches 1012; Indels 828; Gaps 124;

Qy 113 IFOKIKFFDRPELLRVSYWEIYEIT-----DLLCGTQMKP LIITREDVNRNVYA 165
 Db 221 VMORKLEEHREALGRAQVVVDLLQELTSAEQRNQDL---SQQLLEAHSHTLRNTMEA 277
 Qy 166 DLTEEVVYTSENALKWITKGK-----SRHYG-----ETKANQRS 201
 Db 278 ERCESKILMEKVELEMAEREKELYQLQGLEARAQAELEMOYFTLOORHTEEMEEK-- 335
 Qy 202 RSHTIFRMILESREGEPCNCSGVKVSHLNLVDLAGSERRAACTGAAGVRLKEGCNNIRS 261
 Db 336 ---TACISLSLQNEQBELQACD-ALKEENSLLQ-EQEQAAKSAQALOQLEDE----- 384
 Qy 262 LFILGQVIKKLSDQVGGFINTRD-SKULTRIQLNSLGGNPKTRIICTIPVSFDELTLAL 320

Db 385 ---LOQKSISQ-----FVNKNLEKHETSSQSL-----PDVYNEGVQAV 423
 QY 321 QPASTAKYMNTPYNEVSTDEALLAKYRKEIMDLAKOLEEVS-----LETRAQAMEKQ 375
 Db 424 MEESVASLOQR-----VILENEKAGALLSSLEBELRAENEKLSRITLLEAQNAGEADG 480
 QY 376 LA-----QJLEE--KDLQKVQNEKIENTRMLVTSSLT-----LQOEL 413
 Db 481 MVCEVSTAGTALLNRSDSSSTEESSQDVLENTFSQKHKELSVLLVEMKEAQEETAFKLSQ 540
 QY 414 KAKRKR-----RVTWCLGINKMNS-----NVADQF 440
 Db 541 QGKRPEGDYVLDKREYQMMESBGLPSVTARDVLC---APRDKNPSVAVEGEQAGMRDQH 597
 QY 441 NI-----PTNLT-----TKTKHL-----STNLTREI 461
 Db 598 GTLEAGPLNDTGMLNSPPQGDVDSLSAPHVCLCHGELERLKTQVLELTSIHTAKET 657
 QY 462 DESVCSBESDVFSNTDLTSLBIEMNPATKLNQ--ENIESBLNSLRADYDMLV-----LD- 513
 Db 658 HEKNLSEK---AKEISSLTQI---TKFKESAEEARSTLTAVCEERDQLLYRVKELDV 709
 QY 514 YEOLRTEKEMELKKEKNLDLDEPFALEKTKKQOEMQLTHEISNLKMLVKHREVYNODL 573
 Db 710 LGELRAQVRELETSIAE---AEKQGLDYESQRAQHMLLTQIHSLSIEAKSKDVKITL 766
 QY 574 ENELSKVLELLREKEDQIKLOEYIDSKLENIKMDLSYLSIESIEDPKMKQTLFDAETV 633
 Db 767 QRELDGVLOFSEGTQKLSQSQIQTKE-----SEVLEGAERMK----- 806
 QY 634 ALDAKRESAFLRSENLEKEMKELATTYKQOMENDI--OLYQSULEAKKQOMVDLEKELQS 692
 Db 807 -----DISKEMEELSQAALSQKELEIAQMDQLLEKXKQDVET-LQOTTQE 849
 QY 693 AFNEITKLSLIDKVPKDLLCNLEEGKITDLOKELNKEV-----EENEALREEVILLSE 748
 Db 850 KDOQVT-----ELSPSMTKVMQVLENEKFSLGVEIKTLEKQSLLSLR 891
 QY 749 L-----KSLPEVERLRKEIQDKSEELHITSE 776
 Db 892 AEGAKREQVESGABSSPKPHSHSSAEPEPVCKEALQOELEWLKESQRRKRLQAALIS 951
 QY 777 KDKLFSEVYHVESRVQGLEIGTKDPLATTQSNYKSTDOEFQNFQKTLHMDFOKY--- 833
 Db 952 RKELLQKYSKLEELAKVREE--STKDSL--RESEKRELEEDSKN---KDDPEKYGTS 1002
 QY 834 -----KMLENERNQBEIVNLSYEAQKFDSSIGALKATELSYKTOELQEKTRV-- 882
 Db 1003 EWRELEVSRLTISEKEVEGIRDRDLKEKAAAEELQALVORM---TQDLQNKTKQIDL 1059
 QY 883 -QERLNEM-----EOLKEQLENRDSPLQTV 906
 Db 1060 LOEETITENQATTKFITGMDAGDGSVAKETSVSPPRAGGHEHWKPELEGK---IVDL 1116
 QY 907 BREKTLITEKLOOTLEEVKTL---TOEKDDLQLOESIQIERDQLKSDIHTVNNIDTO 963
 Db 1117 EKEKTQLOKQLOEALISRKAILKKAQEKKE--KHLKEELKEQDAYR--HLQEOFDQGSKEN 1173
 QY 964 EOLRNALSLEKHOQETINTLKSISEEV---SRNLHME-----ENTGETK 1005
 Db 1174 ENIRAPRLQLOAKESTDOOLPGTGOEPHSGESLSEGTPEPASEDLHAAQPSHPGETA 1233
 QY 1006 D-----EFOQKVVGID--KKODLEAKNTQTLTADV--KUNEIIEQQKIFSLIOEKNE 1054
 Db 1234 TLQATVSVAQIQOLKEIEVEKEELELKISST--TSELTKKSEVLLLOEQINEQGLEIGN 1292
 QY 1055 LOOMLESVIAEKEQKTLDKLENIENTIENQEBELRLLGDELKQOEIVAEKHAIXKEGE 1114
 Db 1293 LKAASHEAKAHTEQLOEL--ESSQKIALDLEHLKTLQPELETLOKHVGQKEEVSVLYGQ 1351
 QY 1115 LSRTCDRLAEVEKLEKESQ-----OLQEKQOQLANVOEEMSMOKKINIENL 1163
 Db 1352 LGEKEQTLTVQTEMEEEQBRILKAULTQLOEMQAKHEERLKQVQVBEICLKQKPELE--- 1409

QY 1164 KNELKNKELTLEHMETERLELAQKLNENYEEVKSITTKERKVLKELQKSF-----E 1213
 Db 1410 -BESKAOQLOKQLOAALISRKEALKENKSLQOLSSARDADAVEHLTKSLADVESQSVQN 1468
 QY 1214 TERDHLRG-----YIREIATGLQTK-----BELKTAIHTLKEHQTIDELRR 1256
 Db 1469 QEKDALLGKLLALQOBERDKLIVEMDKSLLLENQSLGSCESLKLALGGLTEDKEKL--MKE 1526
 QY 1257 SVSEKTAQIINTODLEKSHTKQOEIPVL-----HEEQELLPN 1294
 Db 1527 LESVRCSTIAESTEWQEKHELOKEVEVLLIQSYENVSNEARIQHVVSVRQEKQEVYAK 1586
 QY 1295 V-----KKVSETQETMNELE----- 1309
 Db 1587 LRSAESDKRERKQLODAEQEMEEMKMKRPAKSKQOKILELEENDRDLRAEAPVPGA 1646
 QY 1310 -----LLTBQSTTKDOSTTLARIEMERLRINEXFQESQEBIKSITKERDNLTIKTKEALEV 1363
 Db 1647 NESMEALLSNAELKEE--LERITLEVKTLSKEFEALMAEKNTLSEETRNLLKQVEAQEL 1704
 QY 1364 KH-----DOLKEHIRE--TLAKTOESQSQOESLNMKEKONETTKIIVSEMEQFPKPD 1413
 Db 1705 KQASLETTEKSDPKDVIEEVTAVVGKSOEODSLSENAKLEDAEATILANSA---XP--- 1759
 QY 1414 SALLRITIEMLGLSKRLQESHDEMKSVAKEKDDLO--RLQEV--LOSESD---QLKENIKE 1467
 Db 1760 -----GWSETF--SSHDDINNVLQDLQKLGRIAELEMEKQKDRLSQLENEKN 1807
 QY 1468 IVAKHLETEB--ELKVAHCCLKEQEBETINELRNLSEKETISTITQKOLEAINDKLQNKIQ 1526
 Db 1808 ALLTOISAKOSELKJL---LEEEVAKINMLNQIOEELSRTVKLTAEAEKDDLEERL- 1862
 QY 1527 EIVEKEBQMLNKQISEVOENVNELKQFKEHRKAKDSALQSTESKMLTETNLOESQEBIQ 1586
 Db 1863 -----MNQLAELNGSIGNTYQDVTDAQIKN---EQLESQMLKRCVSELEEKQ 1909
 QY 1587 IMIKE-----KEEMKRVQEA---LOIERDQKENTKEIVAKMKESQEKYQPLK 1632
 Db 1910 QLVKEKTKVSEIRKRYNEKTOGAQPGPSKIHAKELQELKEKQOEVKQLOKDCIRYLG 1969
 QY 1633 MTAVNETQEKMCE-----TEHLKE---QFETOKLNIENIETENIR 1669
 Db 1970 RISALEKTVKALEFVHTTESQKDLATKGNLAQAVEHHKKAQELSSFXILLDDTQSEAR 2029
 QY 1670 LTQILHNLEEMRSVYTERDDLRVSVEETLKVVERQOLKENLETITRDLKEQELKIVMH 1729
 Db 2030 ---VLADNL-----KLKELQSNKESIK---SQIKOK--DEDLRLRLEQAE-----KH 2070
 QY 1730 LKEHQETIDKLGIIVSEKTNIEISNMQKDLHSNDALKQAQDLKIOBELRIAHMHLKEQOET 1789
 Db 2071 RREKKNQOEKUDALHREKA-----HVEDTL---AEIQVSLTRKDKMKELQOS 2115
 QY 1790 IDKLRGIVSEKTDKLSNMQKOLEN--SNAK-----LOEKIOELKANEHQLITLKD 1838
 Db 2116 LDSTLAQLAATFKSSMSI--QDDRDVRVIDEAKKWEQRFDAIQTKEBEVRLKEENCFTALK-- 2173
 QY 1839 VNETQKVSMEBQKKQIKDQSLTSLKLEIENLNLAEHLHEN--LEEMKSVKBERDLNR 1896
 Db 2174 -DQLRQMTIHEELK-----ITVSRLEHD-----KEIWESKAQTELOHQOQKADYKQ 2220
 QY 1897 VEETLKLERDOLKESLOETKARDLEIQOBLKTARMLSKHEKTEVDKLR-----KI 1947
 Db 2221 ENKELMSQLEAGQIYHSKNETLKLESLKSLKQOSTDLKNSLEKREHNNLEGIKQ 2280
 QY 1948 SEKTIQ-----ISDIQKOLDKSKD---ELQKIQELQKKELOLLRVE-----DVNM 1991
 Db 2281 QEADIQNCFNCEQLETDLTASRELTTRLHDEINVKEQKIIISLLSKEEAAIQVIAIELHQ 2340
 QY 1992 SH--KKINEMEQKQOEPFNPLCKCEMDNFQKTKLHESLEETIRIVAKER-----DELR 2043
 Db 2341 QHSKEITKELENLLSQEEENL--TLSEENKRAVEKTNQLTAELETIKKESLEEQAKLDSEFV 2399

1204 VLKELQSPETER---DHLRGYIREATGLQTKBEL-----KIAHILHKEHQT 1250
 718 KLQEQQSDLEQERRAKEKLQEQQSDLEQ-RAKEKLQEQQSDLEQDLRAKEKLQEQQSD 776
 1251 IDELRSSVEKTAQIINTQOLEK---SHTKLQEEIPVJHEQELLNVKVSSTQETWNE 1307
 777 LEGERRA-KEKLEQ---QOSDLEQDLRAKEKLQEQQSDLEQERRA---KEKLEQEQQSDLEQ 830
 1308 LELLTQSTQSTTTLARIMERILNEXFOESQEBIKSTKERNLKITKEALEVKHQD 1367
 831 ERLAKEKLQEQQSD---LEQER-RAKEKLQEQQSD---BODRL---AKEKLQEQQSD 878
 1368 LKEHIRETLAKIQESQSQSLNMKEK-----DNETTIVSEMEQFKPKDSALLRIE 1421
 879 LEQ-RAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERLAKEKLQEQQSD---LEQ 932
 1422 EMLGLSKRLQESHDENSVAKEDDQLRLQEVLOSE---SDQKENIKEIVAKHLETEE 1478
 933 ERRA-KEKLEQEQQSDLEQERRAKEKLQEQQSDLEQERLAKEKLQEQQSDLEQERLAKEK- 990
 1479 LKVAHCCLEQEQETINELRNLSKEFTEISTIQOLEAINDKLONKIQEIVEKEEQLNIK 1538
 991 -----LOEQQSDLEQERLAKEKLQEQQSDLEQERLA-KEKLEQEQQSD---EQERLAKE 1040
 1539 QISEVQENVNELOKFKHRAKQDSALQSIIESKMLELTLNRLQESQEEIQIMKEK----- 1592
 1041 KLQEQQSDLEQERLAKEKLQEQQSDLE---QERLAKEKLQEQQSDLEQERLAKEKLQEQQSD 1098
 1593 ---BEMKRVQOEALQIERDQKENTKEIVAKMK-ESQEKYQPLKMTAVNETQKWCETIEH- 1648
 1099 DLEQERLAKEKLQEQQSDLEQ---ERLAKEKLQEQQSDLEQERL-AKEKLQEQQSDLEQE 1154
 1649 ---LKEQFETQKINLENTENIRLTQIILHENLEMRSVTKERDDLSVEETLKVERDOLK 1706
 1155 RLAKELQEQQSDLE---QERRAKEKLQEQQSDLEQERTYKASKETILOEQQSDLEQER-LAK 1209
 1707 ENLRETTITRDEKEBELKIVHMHKEHQETIDKLRGIVSEKTNESNNQKOLE---HNSD 1763
 1210 EKLQEQ-OSDLEQERRAK---EKLQEQQSDLEQER-LAKEKLQEQ---QOSDLEQERRAKE 1261
 1764 ALKAQDLKIQBELRIAHMHLKEQSTIDKLRGIVSEKTKLSNMQKDLNSNAKLOEKIQ 1823
 1262 KLQEQQSDLEQERR-AKEKLQEQQSDLEQER---RAKEKLQEQQSDLEQERL-AKEKLQ 1315
 1824 ELKANEHQJITLKDVNETQKVSEWQOLKKQIDQSLTSLKLEIENLNAQELHENLEE 1883
 1316 EQQSDLEQERLAKEKLQEQQSDLEQERRAKEKLQEQQ---SDLEQERL-----AKEKLQEQ 1367
 1884 MKSVMKERDNLRRVEETLK-----LERDOL-KESLQETKARDLEIQELKTARMLSKHEK 1937
 1368 QOSDLEQEQ---RAKEKLQEQQSDLEQERLAKEKLQEQ-QORDLEQERRAKEKLQEQQSDL 1423
 1938 ETVDKLREKISKTQISDIQDKDQSDLEQELQKQELQKLOLRLVKEDVNMHKKIN 1997
 1424 EQERRAKEKLQEQ---QOSDLEQEQ-RAKEKLQEQQSDLEQER---RAKEKLQEQQSD- 1474
 1998 EMEQLKQEPNVLCKCENDMQLTKLHESLEIRIVAKERDLRRIKESLAKMWERDQFI 2057
 1475 EQERLAKE-----KLQEQQSDLEQERRAKEKLQEQQSDLEQERR--- 1513
 2058 ATLREMIARDQNHQVK---PEKRLSDGQOHLMESIREKCSRIKELLKRYSEMDDHYECL 2115
 1514 -----AKEKLQEQQSDLEQERLANEKLQEQQSDLEQEQ-RRAKEKLQEQQSD----- 1557
 2116 NRLSLDLKEHIEPHRTMKLVYVSVTKIKEQHECINKFEMDFIDEVEKOKELIKIQ 2175
 1558 -----DLEQEQ-----RRAK-----EKLQEQQ-----SDLEQERRAKEKLQ 1587
 2176 HLOQDCDVPSERLDLKL-NONMDLHIEETLKQ---FSESEFPISIKTEFOQVLSNRKEMT 2231
 1588 ---EQQSDLEQERLAKEKLQEQQSDLEQERLAKEKLQEQQSDLEQERKADTKNLERKKEHG 1645
 2232 QFLEEWLNTFRDIEKLNKNGIQENDR-----ICQVN---NFFNNRIIATMNESTEF 2280

1646 DVLAEDLYGLRIPALP---LPSENERGYIIPHOSSLPQDNRGNSRDSKEISITIEKT---N 1700
 2281 ERSATISKEWEOQLKS---LKERNEKLFKNYQILKISLASGAVNQVPTQDNKPHVTSRAT 2338
 1701 RESIITNVGRRDIHKGHEKKD-----GSTKPEOKEDKSDIADQNHYL 1744
 2339 Q-LTTEKIRELENSLHEAKESAMHESKILKMKQKLEVTNDIIAKLQAKVHESNCKLEKT 2397
 1745 ETVNISDVNDFOISKYDEDEISAYDSDLDLDE-----EEDDEDLDEF 1785
 2398 KETIQV---LQDKVALGAKPYKEIEDLKMVKIDLEKMKNAKEFEKEISATKATVEYQK 2455
 1786 KPIVQVNDQDENIGI---YK-ELEDL-----TEKNENLDLDEGI-----EK 1825
 2456 EVIRLREMLRRSQOQADTSVISEHTDPQSPKPLTCGGSGIVQNTKALILAKSEHRL 2515
 1826 SSELSEKIKKGKYEKT---KNNFKENDKS-----LYDEHIKKY 1864
 2516 KEISKLKQNEQLIKQ-----KNELLSNNQHLNNEV 2546
 1865 KNDQVNEKEKEFKISLFIHFDGNEILQIVDELSEDI 1902

RESULT 7

T30335
 KLP2 protein - African clawed frog
 C:Species: Xenopus laevis (African clawed frog)
 C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 02-Jun-2000
 C:Accession: T30335
 R:Boletti, H.; Karsenti, E.; Vernos, I.
 Cell 84, 49-59, 1996
 A:Title: Klp2, a new Xenopus centrosomal kinesin-like protein required for centrosome
 A:Reference number: Z20827; MUID:96140639; PMID:8548825
 A:Accession: T30335
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1388 <BOL>
 A:Cross-references: EMBL:X94082; NID:g1129172; PID:e213754; PIDN:CAAG3826.1
 C:Genetics:
 C:Superfamily: unassigned kinesin-related proteins; kinesin motor domain homology

Query Match 7.9%; Score 1048.5; DB 2; Length 1388;
 Best Local Similarity 24.7%; Pred. No. 6.2e-21;
 Matches 424; Conservative 271; Mismatches 561; Indels 463; Gaps 62;

QY 2 AEEGAVAVCVVRPLNSREESL-----GETAQVYKTDNNVI--YQVDSKSFNDFVPHG 55
 DB 22 AEEDAIVFVIRP--PVEGTLTGVDGEGLCLTALSTTIRLHSPKPEPMFTFDHVAHV 79
 QY 56 NETTKNVYEETAAPIIDSAIQYNGTIFAYGOTASGKTYVMGSEDH-----LCVIRP 108
 DB 80 DTQESVFSSVAKNIVESCMNGYNGTIFAYGOTSGKTFMLGPSSEDNFTNLRGVIR 139
 QY 109 AIHDLF---QKIKFPDREFLLRVSYMEIYNETITDLCGTQKMKPLIREDVNRNY 163
 DB 140 SFYFLINRKEKRAKEGKSLCKSFIEIYNEQIFDLL--DSASAGFLREHIKGVF 197
 QY 164 VADLTEVYVYSEMALKWITTKGEKSHYGETKQNSRSRSHIFRMILRESKEGFSNCE 223
 DB 198 VVGAVQVVTSAEAAYQVLSMGWRNRVASTSMNRSRSHAVFTTIESMEK---TNDL 254
 QY 224 GSVKVSMLNLDLAGSERRAAQCAAGVRLKEGKGNINRSLFILQVTKKLSGQVG--GFI 281
 DB 255 VNIRSQLNLDLAGSERQKDTQTEGVRLEKAGSINRSLSCIGQVITALVDVANGORHI 314
 QY 282 NYRDSKLTRLQNSLGNPKTRITCITTPVS--FDETLTALQFASAKYMNTPYNEVS 339
 DB 315 CYRDSKLTFLRDSLGNKATFYIANVHFGSKCFGETSLTQPAKALIKKAVNVE-- 372
 QY 340 TDBALLKRVKEMTDLKKLEFVSLTQPAAMEKDQALQLEEKDLLQKQVNEKIELTR 399

Db 373 -----DTQGVUSQLOAEVKKL-----KEQLSQL----- 396
QY 400 MLVTSSSLTLOELAKRRVVTWCLGKINKMKNSYADOFNPTNITTKHLSINLR 459
Db 397 -----SGQPGDISVARVPSV-----GDNNDWNNFIEAMWILEKSDREKKVLLQ 442
QY 460 EID--RSVCSDESVPSTLTLSEIENWPATKLNQENISELSNLRADYDNLVLDYEQ 517
Db 443 KVVQLEDLCKKKEF-----IQSNKMWKVPREDHIS----- 473
QY 518 RTEKEEMELKAK- EKNDLDEFEALERKTKQDMQMLTHEISNLKULVKHR- EYVNDQLEN 575
Db 474 RLEKAHKEGRISLSNNEODFIA-----ELKEERTLKQVEHHPVAKVALEN 522
QY 576 ELSKVELLEKEKDIKQLQYIDSQ-----KLENIKMDLSYSLESTEDPKMKQTL 627
Db 523 -----HSLREENKRLHLSQSVKRAQEVTAQMALEKAFLEVSUS- EKDRQVAPMSTP 575
QY 628 FDAETVALDAKRESAFLRSNLEKKEKKELATYKQMDNDIQLYQSLEAKKKQMDVLE 687
Db 576 IQLDNNSL---MSAARMREMLQLE---SELATSKOE-----YEEFKELTKKQVEQE 622
QY 688 KEIQSAFNEITTKLTSIDGKVPKDLNLELEGGKITDLOKELNKEVEENALREEVILLIS 747
Db 623 SELQSLIKSNQHLNILEA-----IKANKRHEVSQANRMAHAETI 662
QY 748 ELKSLPSEVERLKEIQDSEELHIITSEKOKLFEVHVHESRVQGLLE- EIGKTKDDL 805
Db 663 KMTTPPTKSYNLSRLVPR-----LSPDAMPNGMLDTPKSGDVMDDI 704
QY 806 ATTQSNYKSTDOEFQNFKTLHMDPEQYKQVUFEENRMQOEIVNLSKEAOKFDSSGALK 865
Db 705 IN-----EPIPPEMS-----EQAYEAIABEELRIVQEV----- 732
QY 866 TELSXYTOBLOKTRVBOERLNEMLKEOLENRDPSLOTVEREKLITEKLOOTLEEVK 925
Db 733 TALQAKLDEEGKNTLQOQVNLKELCSTQIQ-----ELFNRSERNMNMKEQDILIAQIK 786
QY 926 TLTOBKDDLQOLQESLQIERDOLKSDIHDVTNMNIDTQBLNALSLESLKHQDETINTLKS 985
Db 787 SLEKQKQENKS-----QEDVLKSEVHD-----LAVLQOS----- 815
QY 986 KISEVSNLHMEENTGETKDFQKQMGVIDKKQDLEAKNTOTLTPADVK---DNBIIRQQ 1042
Db 816 -----ADRELGAVKGEYSL-----YREKQEKLSQLSARMDVOLQDNDVRLEHE 860
QY 1043 RKIFSLIOEKNELOQM---LESVIA-EKEQLKTDLK---ENIENTIENQSELRLGDEL 1094
Db 861 -----TLLEKESLQDAFNLEEVKFEIDQLKQEIISDSKHENETLRAEFSNLELLETEK 916
QY 1095 KQOEIVAQEKHAIKKEGELSRCTDLRAVEEKLKESQLOKQOQQLLVQVREMSMQ 1154
Db 917 ERROKLTSLQEE-----DKENTKELLQVVDENHWRKOCSELMTKCE 959
QY 1155 KKINEIENKNEKNKELTLEHMETERLELAQKNENYEVKSIKERVKLKQSPET 1214
Db 960 QOVTELHGLEHSLTSKEMIADE-----KONTADKEVVADLMQNIQV 1002
QY 1215 ERDLRGVIREIATGLQTKELKIAHIH-----LKEHQETIDELRSV 1258
Db 1003 HRTTI---IHKTESIDLLTRE---LEDIHSKYSIVLLAKEESKTVIEEQEKQIBELRECL 1056
QY 1259 SEK-TAQIIN-----TODLESKHTKLOEIPVLH-EEOELLPNVKVSSETORTM 1305
Db 1057 ERKQSAADNIEKELCCDLAHATELEKLTAFNQEALHTHEKEL---VEKQOQISELT 1113
QY 1306 NELELLTEQSTTKDSTTLARIEMERLRLNEKFBQS-----QEBIKSLTK 1349
Db 1114 NOVKLMTDLFISR-----EQEKIRPASSNSSPPVLPETPRTPGPNFVDSIANLQK 1165
QY 1350 ERDNLTKIKALEVKHDLQKHEHRETILAKTQESQSKOEQSLNKEKKNETTKIVSEMEQF 1409
Db 1166 RNTNLEILVSELN-----BERTSKNEEIRLKMQLCETEN----- 1200

QY 1410 KPXDALLRIEIML-GLSKRLQESHDEMSVAKBKDDQLQRLQVLOSQDQKLEIKI 1468
Db 1201 -----MRLEIQNLQMGCKELKSQLENCNVMKDSNDQK-----PSDMQDLAKREIEKE 1247
QY 1469 VAKHLE---TEBELKVAHCCLKEQETINEIRLVNLSSEKTEISTIOKLEAINDKLQNK 1524
Db 1248 VSRMEKKGKATEHILKLQ---AELEETRNL-----CTKXHSINE 1284
QY 1525 IQBIYKEBEQNLTKOISEOVENVNLEKQFKEHKKAKDSALQSISTESKMLELTNR-----L 1578
Db 1285 LSKEIERTSLRAKAFERKEEIRSEILLEGKYEEYEKLSHELDMLRKQVFLAEENGKILGH 1344
QY 1579 QESQBEIQIMIK-EKEMKRVQEA--LQIERDOLKENTK 1614
Db 1345 QNPQKIQIYLVKLKKNENKLEBAEKLRLENLFLKESKK 1383

RESULT 8
T08621
centroome associated protein CEP250 - human
C:Species: Homo sapiens (man)
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: T08621
R:MacK, G.J.; Rees, J.; Sandblom, O.; Balczon, R.; Fritzler, M.J.; Rattner, J.B.
A:Title: Autoantibodies to a group of centrosomal proteins in human autoimmune sera react with CEP250
A:Reference number: Z16462; MUID:98165428; PMID:9506584
A:Accession: T08621
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2442 <MAC>
A:Cross-references: EMBL:AF022655; NID:g2832236; PIDN:AAC06349.1; PID:g2832237
A:Experimental source: cell line HeLa

Query Match 7.7%; Score 1026; DB 2; Length 2442;
Best Local Similarity 20.5%; Pred. No. 4.2e-20;
Matches 555; Conservative 561; Mismatches 945; Indels 642; Gaps 109;
279 GFINYRDSKLRILQNSLGGNPKTRIICTITPVSPDELTALQFASTAKYKNTYVNE- 337
7 GLNNMKPQSLQVLEEQV-----LALQQQMAENQAASWRK-LKNSQEAQOR 51
338 -----VSTDEALLKRYRKEIMDKKQL-----EVSLETQRAQAMEKQD--- 375
52 QATLVKRLQAKVLQVRSWCQLEKRLKLEATGPGIPQWENVEPNLDELLVRLSEEQRCE 111
376 -LAOL-----LEEKOLLQKVQNEKIENTL-----RMLVTSSSLTLOE-----LK 414
112 SLAEVNTQIRLHMEKADVNNKALRADVEKLTVDWNSRARDLEMRKESQWMEQEFKGYLK 171
415 AKRRRVTWCLGKIN-----KMKNSVADQFNPTNITTKTH-KLSINLRLDESVC 467
172 GEHGLLSLREVVTFRRHFLEMSATDRDLMELKAE-----HVRLSGSL-----TCCL 221
468 ESDVFSNTLTLSEIENWPATKLNQENISELSNLRADYDNLVLDYEQRLTEKEEMELK 527
222 RLTVGAQR-----EPNGSRMDGRE--PAQLLLLLAKTQEL---EKEAHERSQELIQ 269
528 LKEKNDLDEFEALERKTKQDMQMLTHEISNLKULVKHREYVNDQLENLSKVELLEK 587
270 LKSGDLEKALQDR-----VTLSALLTQSQKONEDYERKMINA----- 308
588 EDQIKKIQEYIDSKLENIKMDLSYSLESTEDPKMKQTLFDAFTVAL---DAKRESAFL 644
309 ---LRETVEILETNHTELMHEASLSRNAQEKLSQGVTKDITQNMVBERGDNIAQSGSL 365
645 RSNLELEKQKELATYKQMDNDIQLYQSLEAKKKQMDVLEKELQSAFNETKLTSLI 704
366 EN-SLESSIFS-QPDYQADAKALTVRSVLTRRQAVQDLQQLAGCQAEVNLQOQH 423
705 D-----GKVPKDLICNL-----ELEKIDTLOKE---LNKEVEENALREEVILLSELKS 751

Db 424 DOWEEGKALRQRLQKLTGERDTLAGOTVDLQGEVDSLSKERELLQKABEE--LRQQLLEV 481
Qy 752 LPSEVERLRK---EQ-----DKSEELHIITSEKDKLFSEVHVHKEGRVQGLLEE 797
Db 482 LEQEAWLRLRVNVELQLOGDSAQKQEOOEELHIAVRERERLOEMLGLEAKQESLSE 541
Qy 798 IGTKDDLTATQSNKYSTQEQFQNFXTLHMDPEQKYMVLEENRMOBIVNLSKEAQKF 857
Db 542 LITLREALSEIHLLEGELLQEQTEVTAALARAQSIATLSSSENTEKTEVADLRAAAVKL 601
Qy 858 DSSIGAL---KTELSYKTQELQKTRVQERLNMELQKLEENRSPLOTVEREKLIT 914
Db 602 SALNEALALDKVGLNQLOLLEENQSVCSRMAEABEQARNALQ---VDLAAEKREALW 658
Qy 915 EK---LOOTLEEVKTLTOE-KDLKQLOESLQIERDQLKSDIHDVTNMNIDTQOELRNAL 970
Db 659 EKNTHLEAQLOKABEAGAELOALDRIOER-----KEEIOKLSSESHQOEATQOL 710
Qy 971 ESLKQHOETINTLKSISERB-SRNLHMEENTGETKDBFOQKMGVIGDK-KODLEAKNTQT 1028
Db 711 EQL--HOE-----AKQOEVLARAVQKEALVREKAALVRLQAVERDRQDLAAQ-LOG 761
Qy 1029 LTA-DVKONEIIEQQRKIFSLIQKNELOOMLESVIAEKEOLKTDLKENIEMTIOE 1086
Db 762 LSSAKELLESSLFEAQOONSVIDEPOGQLEVOIQVTQAEVI-----QGE 807
Qy 1087 LRLIGDELKQOEIVAOEKHAIKKEGELSRTCDRLAEVVEKLEKESQQLQEQOQLNV 1146
Db 808 VRCLKELDTERSQAEQERDAAAAR-----QLAQAEQGTALQEQKAAHEKVNQ 857
Qy 1147 QEEMSEMOKKINETENLKNELKNKELTLEHMETERLELAQKLNENYBEVKSITKERVULK 1206
Db 858 LREKWEKERSHQELAK-----ALESLEKWELEMLKBEQOTEMEALQAEER 909
Qy 1207 E-----LOKSFETERHLRGYIEIEATGLQTEKELKIAHILKEHQETIDELRRSVSE 1260
Db 910 TOAESALCOMQLETEKERV-----SLLETLLQOTKELADA-----SQOLERL 952
Qy 1261 KTAQIINTDLEKSHTKLOEIPVLHBEQELLPNVKVSETQETWNELELLTEQ----- 1314
Db 953 -----QDMKVQKLEHQTGILQ-----TQLOEAQRELKEAARQHRDLA 992
Qy 1315 STTKDSTTLARIEMERLNEKFORSEIEKSLTKERNLNTIKIEALEVKHDOLKEHIRE 1374
Db 993 ALQESSLLQKMD-----LOKQVEDLSQVLAQDDSQLVE-----QEVOEKLE 1039
Qy 1375 T--LAKIQESQKQESQSLMKEKNETTKIV-SEMEQFKPKDSALLRIEIMGLSKRLQ 1431
Db 1040 TOEYNRIOKLELEREKASITLSIMEKEBQRLVLQOEAADSIRQOELSALR-----QDMQ 1090
Qy 1432 ESHDEKMSVAKEKODLQRLQEVLOSQESDOLKEN---IKEIVAKHLETEBELKVAHCCLKE 1488
Db 1091 EAQGEKLSAQMEILR--QEVKEADFLAQEAQLLEELASHI-TEQOLRAS---LWA 1144
Qy 1489 QETINELRVLSEKTEISTI-----QKLEAINDKLONKIOEYKEBQI- 1535
Db 1145 QEAQAQLHLRLARSTESQLEALAAEQPGNQAQAQALASLYSALQALGSCVCESRPELS 1204
Qy 1536 -----NITKQISEVQENNELKQFKEHAKDQKADKALOSIESKMLELTNQLQESQEIQIM 1588
Db 1205 GGGDSAPSVGLEPQNGARSL--FKRGPLLTALSAEASVALKHLQDLQKMTQOTRDVL 1262
Qy 1589 IKEKEEMKVOEAL---QIERDQKENTKEIVAKKESQEKYQFLEKMTAVNETORCMCE 1645
Db 1263 ---RDQVQKLEERLDTAEKKSQVHTELQDLQRLQSLQONQBEKSKW-----EGKQNSLE 1312
Qy 1646 IEHLKEQFET-----QKLANLEIETENIR-LTQILHLENL---EEMRSVTKE----- 1687
Db 1313 SE-LMELHETMASLQSLRLARAELOEAMEAQGERELLQAAKENLTAQVEHLQAAVVEARAQA 1371
Qy 1688 -----RDLRSVEETLKVREDQKLENRETTITRDEKOBELKIVH-----MH 1729
Db 1372 SAAGILEEDLRTARSALKLNKEEV-ESERER-AQALQEQGELKVAQKALQENLALLTQT 1429

Qy 1730 LXEHOETIDKLRGIVSEKTNETISNMOK-----DLEHSDALKAQADLKIOE----- 1774
Db 1430 LAAREEEVETLRGQIOLEKQ-REMQKAALELLSLLDKKNOEVDLQEQFOIOELEKCRSV 1488
Qy 1775 -----ELRIAHMLKQEQETIDKLRGIVS-----EKTDKLSNMOK-----DL 1811
Db 1489 LEHLPMAVQERQKLTVQREQIREPEKDRQETQNVLEHQHLELEKQDOMIESQORGQVDL 1548
Qy 1812 ENSNAKLOKIOELKANEHQLITLKKOVNETQ-----KKVS-----EMEOLKQIKDOOS 1860
Db 1549 KXQVLTLECALALEENHMKECQKLIKLEBQORETQVALTHLTLIDLEERSQEIQAQOS 1608
Qy 1861 LITLKLIEINLNAELHENLEMKSVKERNLRRVEE--TLKLERDQKLEOFTKAR 1918
Db 1609 SQIHDLESHSTVLARELQERDQVKSQREQIELOKQKEHLTQDLERRDQELMLQKERIQ 1668
Qy 1919 DLEIIOELQELTARMLSKEHETVDKLRREKISEKTIQISDIQKDLKSKDELOKQIOELQKK 1978
Db 1669 VLEDQRTROT-KILEEDLEQIKLSLRERGRELTTQRLMQERABEGKPSKAQGRSLEHM 1727
Qy 1979 ELQOLLRVKEDVNMHSHKINEMEQKQFEPNYLCKCEMDNFQTLTKLHESLEIRIV-- 2035
Db 1728 KLILRDKQEKVECCQEHIELQELKDOLE-----QQLQGLHRKVGETSILLSQ 1775
Qy 2036 -----AKERDELRRIKESLSKMERQFIATLEMIARDRQNHQVPEKLLS 2081
Db 1776 REOEIVVLQOOLQERAEQELK--EQSLOSQLEDE--AQALAQRDQLEALAQEQEQ-QA 1829
Qy 2082 DGOQHLMESLRKCSRIK-----ELLKRYSEMDDHYECLNRLSLDLKEKEFHRIMK 2133
Db 1930 QGQE--ERVKEKADALQGALEQAHTLKERHGBELQDHKEQARLEBELAVE--GRVQ 1883
Qy 2134 KLYVLSYVTKIKEBQHECINKEMDFIDEVQK-----KELLIKIHLQOQCDVPSR 2186
Db 1884 ALEVLGDLRAESREQKALLAQOQCAEQAEHEVETRALQDQSWLQAOVLKRED---Q 1940
Qy 2187 ELRDLKLNQNDLHIEILKDFSESSEFPKSTEFQVLSNRKEMTQFLEEWLNTFRDIEK 2246
Db 1941 ELEALRAESQSSRHQEAARABEA-----LQEALGKAHAALQGEQHLLEQAELSR 1992
Qy 2247 LKNG-----IQENDRIQVNNFNNRIIANNESFEERSATISKWEODLSKKEKNE 2302
Db 1993 SLEASTATLOASLD-ACQASHRQLEEARLI--QEGEIQDQDLRYQEDVQOQOALAQARDE 2049
Qy 2303 KL---FKNYQTLKTSLSGAQVNETQDNKNPHVTSRATQTLTKERLENSLHAKESA 2359
Db 2050 ELRHQEREQOLLEKSLAQRVQEN-MIOEKQN-----LGLEREEIEIRGHQSVRELQTL 2103
Qy 2360 MHEKSKITKMOKELEVTNDIITAKIQAQVHESNKC--LEKTKETIQVLOQDKVALGAKPYKE 2417
Db 2104 AQKEOEIIELEKRTQORNN-----LEALPH-SHKTSPEEQSLKLDLSLEPLR-----OR 2150
Qy 2418 EIEDLKMVLKI---DLEKMKNAKEFEKEISATKATVYQKEVIRLLR-----ENLRSSQ 2470
Db 2151 ELERLOAALRQTEAREIEWREKAQDLALSQAQTKASVSSLOEVAMFLQASVLERSEHOR 2210
Qy 2471 AQDTSVISE-----HTDPOPSNKPLTCGGSGIV-----QNTKALIL 2507
Db 2211 LODELELTRALKEKRLHSPGATSTAELGSRGEQVQJGEVSVABESPDMGKQSWRQ 2270
Qy 2508 KSEHI-----RLEKISKLKQONEQIIOKNEILLNNQHLNNEVKTWKERTLKREAHK-- 2560
Db 2271 RUEHLQOAVARLEIDRSRLQRHVQ-----LRSTLEQVERE-----RRLKREAMRAA 2318
Qy 2561 -----QVTCENSPKSPKVTGTA 2577
Db 2319 QAGSLEISKATASSPTQODGRGKNSNAKCVAELOKEVVLLQAOQLTLERKQODYITRSA 2378
Qy 2578 SKKKQI-----TPQCKERNLQDPVPKESPKSCFSDRSKSLPSP 2617
Db 2379 QTSRELAGLHLSHLSLAVAPAEATVLEAETRLD-----ESITQSUTSP 2425

[illegible]

Db 301 RUSQVGVGFKDLQWQVSDSKYARLLKLGIRHQSVAFKLNASSRSHSIPVTKIL 360
QY 213 SREKGEPSNCEGSKVSHNLVDLAGSERAATGAGVRLKGGCINRSLFILGVQIKL 272
Db 361 QIEDSEMSRV---IRVSELCDLAGSRTMTQNEGERLRTGINTSLLTGLKCIYVL 417
QY 273 SGQVGGF---INRDSKLRILONGNPNKTRITCTTPVS---FDELTALQFAS 324
Db 418 KNSKSKFKQHPFRESKTHYFQSPFGKKG---ICMIVNSQCVLAYDETLNVLFKFA 474
QY 325 TAKYMKNTPVYNEVSTDEALLKRYRKEIMDLKQLEEVSLT-----RAQAMEKD 374
Db 475 IAO-----KVCVPDT--LNSQDKLFGPVKSSQDVSLSNSKILNVKATISWEN 524
QY 375 QLAQLLEKDLQKQVON--EKIENL--TRMLVTSSLSLQOE---LXAKRRERVTCWLGKIN 429
Db 525 SLEDLMEDEDLVELENAEETQNVETKLLDEDLDTLEENKAFISHEEKKLLDLIEDLK 584
QY 430 KM-----KSNYADQFNIPNTITTKHKLINLLREIDESCSDVFSNLTDLTSELEWN 485
Db 585 KCLINEKKEKLTLEFKLIREVTOFTQYWAQRAADPKETLLQREILEENAEARLAFKD 644
QY 486 PATKLLNQNIESEILNSLRADYON---LVLDEYQLRTE-----KEEMELKUKEN 532
Db 645 LVGKCDTREAAKDI CATKVTEBEATACLEKFNQIKAEAKTKGELIKTKBELKKRENE 704
QY 533 DLDFEALERTKK-----DOEQMLIHEISNLKN-----561
Db 705 SDSLIQLETSNKKIITQONRIKELINIIDQKEDTINEFONLKSHMENTFKCNDAKATSS 764
QY 562 LVKHRVYQDLENELSSKVEL-----LREKEDQIKKLEQYI-----DSQKL 603
Db 765 LIINNKLICNETVEVPDQSKSKICSEKRVNENELQODEPPAKKGSIHVSSAITDQKS 824
QY 604 ENIKMDLSYLESIEDPKQKQTLFDAAETVALDAKRESAFARSENLELKEKOKELATYK 663
Db 825 EEVFPNTA---ETEDIRVLOE-----NNEGLRAFLLTIENELKNEKEKAEINK 870
QY 664 QMENDIOLYQQLBAKKQVQVLDLKELOSAFNEITKLTSLIDGKVPKDLLNLEGGKIT 723
Db 871 QIVH---FOELSLSEKKNLTLSKEVO-----QIQSNYDIATAEL 907
QY 724 DLQELKNEVEENAEALREEVILLSELKSLPSEVERLKEIOQKSEELHIITSEKDKLFS- 782
Db 908 HVQSKNOQEZE---KIMLSNEIETATRSITNNVSQIKLMTKIDELRTL 955
QY 783 EVVHKESRVOGL-LEEI--GRTKODLATTSQNYKSTDOEFONFKTLHMDFFQKYKMWLEE 839
Db 956 DSVQISNIDLLNRLDLSNGSEEDNLPNTQLDLLGND-----YLVSKQVKEYRTQE 1006
QY 840 NERMNQEIWNLSKEAKQFDSSLGALKTELSTYKTOLEKTRVQBRNEMELKEOLENR 899
Db 1007 PNREN-----SPHSSIEAT-----WEECKBIVKASSKSHQIIELEQOIEKL 1048
QY 900 DSPLOTVEREKLITELKLOOTLEEVKTLTQEKDDL-KOLOESLQIERDQLKSDIHDVTVM 958
Db 1049 QAEVKGVDENNRKKEHKHQDD---LLKKEETLIQLKEELQ-----EK 1091
QY 959 NIDTQOELRNAL---SLKQHOETINTLKSTSEEVSRNLHMEENTGTYKDFQKMGVI 1015
Db 1092 NVTLDOVQIHVVGKRALSELTOGVTCYKAKIKE-----LETI 1129
QY 1016 DKKQDLKAKNTQTLTADVKDNEIIEQORXIFSLIQEKNELOQMLESVIAEKQL---KTD 1072
Db 1130 LETQKVERSHSAK-----EQDILEKSEITILKERNLKEFQHLQDSVKNVTKDLNVKELK 1184
QY 1073 LKENIEMTTENOBELRLGLDELKQOEIVAOBKHNAIKKE-----GELSRCTD 1120
Db 1185 LKEETITQTLNLODMKHL-LQLKEEBEETNRQETKLEELSSASSARTQMLKADLQKEE 1243
QY 1121 RLAEVEEKLKESQLOEQKQQLNVOEEMSEMKKINEIENKN-----ELKNKELTLEH 1176

1244 DYADLKEKLTDAKKQIKQVQKEVSVNRDEDKLLRIKINELEKKQNCQSOELDMKORTIOO 1303
QY 1177 METERLELAQKUNE---NYEEV-KSITKERVVLKEL-----OKSFETERDH-LRGYIRE 1225
Db 1304 LK-EQLN-NQKVEAIQYERACKDLNVKEKIIEDMRMTLEBQEQOTQVQDDQVLEAKLE 1361
QY 1226 IE--ATGLQOT-KE-----ELKTAHILKHEQHTIDELRRSVSEKTAQIINTQD-LEKSH 1276
Db 1362 VERLATELEKKEKCNLDLETQNNQSRNKEHENTDVL-----GKLTNLQDELQESEQ 1413
QY 1277 K-----LOBEIPVLHE--EQELLPN--VKVSVETQ-----TMELELLTQOSTTKDS 1320
Db 1414 KYNADKKWLEKMWMLITQAKAEANIRNKEMKYAEDRERFFKQOQEMELITLAQLTEKDS 1473
QY 1321 TTLARIEMERLRLNEKFOESQEBEISLTYKERNLKITIKALEVKKHQDLKEHIRETLAKIQ 1380
Db 1474 -----DLQKREERDQVLA---ALEI---OLK-----ALLS 1498
QY 1381 ESQSKQEQSLNKKEDNETTKIVSEMEQKPKDSALL---RIRIEMGLSKRLQESHDEM 1437
Db 1499 SNVOKDNEIQKRIISETSKITQIETQIMDKPKRISSADPKLQTEPLSTSEISRNKIED 1558
QY 1438 KS-----VAKEKDDLQK-----LQEVLOSES 1458
Db 1559 GSVVLDSCEVSTENDQSTRFPKPELEIQFTPLQPNKMAVHPGCTTPVTVEIPKARKRS 1618
QY 1459 DQKKE-----NIKEIVA--KHLETEBELKVAHCCCLKEQETINELR--- 1497
Db 1619 NEMEEIDLKCNKKNATPRTNLKFPISDDRRSNVSKQKQVA---IRSSKKTYSLRQAS 1675
QY 1498 ---VNLSEKETEISTQK---OLEAINDKLQNKIQEIEYEKEEQINIKOISEVQENNELK 1551
Db 1676 IIGVNLATKKKE-GTLQKFGDFLOHSPSILQSKAKKIETMSSSKLSNVEASKENVSQPK 1734
QY 1552 QFKEHRKADKALQS---IESKMLELNLNLOESQEBEIQIIMKEEEMK 1596
Db 1735 RAK--RKLYTSEISSPIDISQVILMDQMKESDHQI---IKRLRTK 1777

RESULT 11
D96619
Protein T30E16.9 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
R:Accession: D96619
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: D96619
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-888 <STO>
A:Cross-references: GB:AE005173; NID:g8778739; PIDN:AAF79747.1; GSPDB:GN00141
C:Genetics:
A:Gene: T30E16.9
A:Map position: 1

Query Match 7.3%; Score 971.5; DB 2; Length 888;
Best Local Similarity 31.2%; Pred. No. 4.2e-19;
Matches 301; Conservative 149; Mismatches 305; Indels 209; Gaps 30;
QY 7 VAVCVVRPLNRSRELSGETAQVYKTDNNVI-----Y 39
Db 4 ICVAVRVRP-----PAPENGASLWKVEDNRISLHKSLDTPITTASHAFVSGISISTDLI 57

[illegible]

RESULT 12
A54803

microtubule-associated motor KIF4 - mouse
N/Alternate names: kinesin-related protein KIF4
C/Species: Mus musculus (house mouse)
C/Date: 05-Apr-1995 #sequence_revision 05-Apr-1995
C/Accession: A54803; D44259
R/Sekine, Y.; Okada, Y.; Noda, Y.; Kondo, S.; Aizawa, H.

J. Cell Biol. 127, 187-201, 1994
A>Title: A novel microtubule-based motor protein (XIF4) for organelle transports, whose
A/Reference number: A54803; MUID:95014709; PMID:7929562
A/Accession: A54803
A/Molecule type: mRNA
A/Residues: 1-1231 <SEK>
A/Cross-references: GB:D126646; NID:G563772; PID:BAA02167.1; PID:d1002657; PID:G563773
R/Aizawa, H.; Sekine, Y.; Takemura, R.; Zhang, Z.; Nangaku, M.; Hirokawa, N.
J. Cell Biol. 119, 1287-1296, 1992
A>Title: Kinesin family in murine central nervous system.
A/Reference number: A44359; MUID:9307686; PMID:1447303
A/Accession: D44259
A/Status: nucleic acid sequence not shown; not compared with conceptual translation
A/Molecule type: mRNA
A/Residues: 91-111, 'S', 113-240 <AIZ>
A/Experimental source: brain
A/Note: sequence extracted from NCBI backbone (NCBIP:118904)
C/Superfamily: unassigned kinesin-related proteins; kinesin motor domain homology
K/Keywords: ATP; nucleotide binding; P-loop
F/180-343/Domain: kinesin motor domain homology <KMO>
F/88-95/Region: nucleotide-binding motif A (P-loop)

Query Match 7.1%; Score 952.5; DB 2; Length 1231;
Best Local Similarity 26.0%; Pred. No. 1.8e-18;
Matches 367; Conservative 223; Mismatches 491; Indels 329; Gaps 51

QY	7	VAVCVRVRLNPGREESLG-ETAVQVYKMTDNNVIYOVDGSKSFNDFRVFPHGNETTKNVYEE	65
Db	10	VRVALRCRLVSKKEGQCTCLSPVGPQVV--VGNDKSYTFYDFVPDSTQEVEFVT	67
QY	66	IAAPIIDSAIQYNGTIFAYGOTAGSKTYMMGS-----EDH---LGVIPRAIHDFQIK	118
Db	68	AVAPILKGVFGYNATVLAIGYGTGSKTYSMGCAVTAEOHDSAIGVPRVIQLLPKEIN	127
QY	119	KFPDREFLLRVSYMEIYNETTLLCGT-QKMKPLIREDVNRNVYVADLTBEVYVTSYM	177
Db	128	KSDREFTLKVSYLEIYNEEILDLCSSREKATQINIREDPKEGKIVGLTEKTTVLVASD	187
QY	178	ALKWTTKGEKSRHYGETMKNORSRSHITFRMILESREKGEPSNCEGSKVSHLNLVDA	237
Db	188	TVSCLEQGNNSKTVASTAMNQSSESHAIFVISIQRYK----NDKNSFSRKLHLVDA	243
QY	238	GSERAAQTGAAGVRLKEGCNINRSLFIUGVVKKLSGQVGGFINRYRDSKLTIRLQNSIG	297
Db	244	GSEHQKTKAEGDRLREGINNRGLLCLGNVISALGDGKGNFVPRYRDSKLTIRLQDSIG	303
QY	298	GNPKTRITCTTPV--SFEFTLTALOPASTAKYMKNTPYVNEVSTDEALLKRYRKEIMDL	355
Db	304	GNSHTLMTACVSPADSNLEETNIRYADRARKTKNRPIIN-----IDPQAAELNHL	355
QY	356	KQOLEEVS-----ETRAQAMEKDQLAQLLEKOLLOKVONEKLEN	396
Db	356	KQOQOQLQILLQAHGGTLPGDINVEPSENIQSLMEXNQ--SLVNEEKLSRGISEAAGQ	413
QY	397	LTRML---VTSSSLTLOELKAKKRRVTCWLGKINK-----MKNSYADQFNIPNTITTK	449
Db	414	TAQMLERIILTEQANEKONAKLEIRRHAAKVDQLKVLTEBQELKENTEIICNLQOV	473
QY	450	THKLS-----INLLREIDSVCSS-----ESDVFNSTLDTLSE-----IENWP	486
Db	474	IAQLSDBAACMTATITAGEADTVQVSSPDTSRSSDVFS--TOHALRQAQMSKELIELNK	532
QY	487	A-----TKLJNQENIESELNIRADYD---NLVDYEQLRTEKEMEELKLA-EKNDL	534
Db	533	ALALKEALAKKTQN--DNQLQPIQFOYQDNINKNLESEVLSLQREKELVLELOTAKKDA	590
QY	535	DEFEALERTKKDOEWQLIHEINLKNLVKHRREVYNQDLENEISSKVLREKEDQIKKL	594
Db	591	NQAKLSERRRKLQELE--GQIADLKKGLQ-----EQSKLLKKESTEHTVSKL	637
QY	595	QEYIDSOKLENIKWDLSYSLESIEDPKMQKTLFDFAETVAL---DAKRESAFURSE-NLE	650
Db	638	NQHIRMKKQORVOL--WRQMKEDAKEPFRMQQK--DKREVTLKERDRKQYELLKJERNFO	695

QY	651	-----LKEKMLATYKQWENDIOLYQSLAEAKKQVVDLEKELQSAFNEITKLTSLID	705
Db	696	KQSNVLRRTTEAAAAANKKALQ-----KQVAAKKEKTOGRGWEASTAARM---	744
QY	706	GKVPKLLCNLEEGKITDLQKELNEVEENALREEVILLSELKSLPSEVERLRKBIOD	765
Db	745	-----KWLGN-----EIFVWSTEEAKRHGLLEERKILQADVAQL-KEKRE	787
QY	766	KSELHIITSEKOKLPSEVVHKSRRVQGLLEEIGTKYKDDLATTSQSNKSTQDFQNFKL	825
Db	788	SGENPPLKLRRTFSYDEIHQDS---GABDSIAK-----	819
QY	826	HMFEOKYKVLNEENRMOEIVNLSKEAQFDSLSGALKTSLSYKTOELOEKTRVQER	885
Db	820	-----QIESLETELELSAQI-----ADLQOK	841
QY	886	LNMEQOLKEQENRDSPLQTVREKTLITEKLOOTLEEVKTLTQEKDDLKQLOESQIER	945
Db	842	LLD-----AESDRPKQWESATILEAKC-----AIKYLVELVSSKILVSKLESSL	889
QY	946	DQKSDIHDVTNNIDTQOLRNALESK-----QHOETINTLSKISSEEVSRNLHM	997
Db	890	NQSKASCIDYQKMLFEEQNHFAKIETELKEELVKVEQOHOEKVLYLLLSQSQSOMTEKQL	949
QY	998	BENTGETKDFQOKWVGIDKKQDLKAEKNTOTLTADVKNELIEOQRKIFSLIQEKNELOQ	1057
Db	950	EESVSEK-----EQULLSTLKQOEEELRMQ-----EVCEQNOQ---LLQENSAIKQ	993
QY	1058	MLESV-LAEKEQLTKLKENIEMTIENO-----BELRLGLDELKKQ	1098
Db	994	KLTLQVASKQ---KPHLTRNIFQSPDSFVEIPPKPKRIKCEKLEQSFAGVGLQYSE	1051
QY	1099	EIVAQEKNAHAKGSELRTCDRLAEVBEKLE-----KSQLOEKQOQLLNQOEE	1149
Db	1052	PSVAEQDNEDGDDHADEWIPTKLVKSKSIQGCSCXGWCNQCQCGRKQKSDCNVSCS	1111
QY	1150	MSEQKKNELNENLKNELKNELTLEHMETERLELAQKLNENY-----EVKS-----	1197
Db	1112	CDPTK-----CNRHONQD-----NSDAIELNQDSNSFKLEDTEVTSGLSFPH	1157
QY	1198	ITKERVKLKL-----QKSFETERDHLRGYIREIATGLQTKKEELKIAHIHKE	1246
Db	1158	ICATPSSKILKEMCDADQVQLQPMFVSSDH-----PELKSTIASSESQ	1201
QY	1247	HOETIDELRRSVSEKTAQIINTQDL-EXSH	1275
Db	1202	NKAIGKKKKRALASNTSPFSGCSPIQEESH	1231

RESULT 13
T18372
repeat organellar protein - Plasmodium chabaudi
C:Species: Plasmodium chabaudi
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18372
R:Werner, E.B.; Taylor, W.R.; Holder, A.A.
Mol. Biochem. Parasitol. 94, 185-196, 1998
A:Title: A Plasmodium chabaudi protein contains a repetitive region with a predicted spe
A:Reference number: Z18922; MUID:98418765; PMID:9747969
A:Accession: T18372
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1939 <WER>
A:Cross-references: EMBL:U43145; MID:g1151157; PID:g1151158; PIDN:AAC63403.1

Query Match 7.1%; Score 946; DB 2; Length 1939;
Best local Similarity 22.1%; Pred. No. 4.2e-18;
Matches 492; Conservative 438; Mismatches 768; Indels 532; Gaps 105;

QY 542 RYTKQO-----EMQLIHEISNLKVLKRVNVDLELSEKVELLREKEQIKQLQE 596
Db 6 KSKKNBDGSKDSKKTNETSGIEKESKNKYNKVNNS-----TKKDKNDNS 56

QY	1597	RVOEALQIBRDOLKENTKEIIVAKMSESQEKYQFLK--MTAVNETOBK-----MCEIE 1647
Db	999	GIEEKYKVAIKLAEBHKDQVTVKLGQEHKEIAKLEDGHGEVNEVEKQASLANMLLEEN 1058
QY	1648	HLKEQFETOKNLNIENTENITLTOILHENLEBWSVTYKDERDDLSRVSVEETLKVERDOLKE 1707
Db	1059	HKNEMI--KLKEEHKESAS-DLVEKLYQKDEEYKNSNNKIETLTNTVIKOINDSIMCYKK 1114
QY	1708	NLRETITRDLKQEE--LKIIVHMLKHEQETIDKLRIGIVSEKNEISNQKOLEHENDA 1764
Db	1115	QILEEVEKHEVNEEINKLIVQEMKDN--DKK-----ILEKNEIILKCNLKCLSN----- 1164
QY	1765	LKAQDLKIQEELRIAPHLKXQOETIDKLRIGIVSEKTDKLSNMKOLENSNAKLOEKIQE 1824
Db	1165	YKVPETK-ENTYKSEWVWNEKER-----LIIVDSVCKENISSESDVEGKGNLKWTL-S 1216
QY	1825	LKANEHQLITLKKOVNETQKV-----SEMQLKKQIKQOGLTSLKLEIENLMAQE 1876
Db	1217	LKKKERNITFSINDKNSESELVDITKSAYINKIEIMYKKEIEDNGKNTEDLKNKILDLSNE 1276
QY	1877	LHENLEEMKSVYKERNDLRRVEETLKLEROOLKESLOETKARDLEIQOELKTAMLSKEH 1936
Db	1277	L-INLENNKVLTDENN-----NLK-----KETEIKD----- 1302
QY	1937	KETVDKLEKISEKTIQISDIQKOLDKSKOBLQKKIQELOKKEQLQLLRVKEDVNMGHKKI 1996
Db	1303	-----NKLKEKKNTEILNLDNDIIK-----LKKEISEWKDEEKL--TKENIKLK----- 1348
QY	1997	NEMQLKKQFPNYLCKCEMDNFOLTKYKLHSLSEETIRIVAKERDELRRIK-ESLKMERDQ 2055
Db	1349	NDIEQINKEY-----KIKEENLMI--KFENINENVTSL-KNGIETEKMKLEELNKNYEL 1399
QY	2056	FIATLRE--MIARDQNHQVKEPKRLASD--GOHLMESLREKCSIRIKELLKRYSEMDD 2110
Db	1400	LLAEKRETNWISINDDKIV--ENNILLETDSQNNLNKNVEDKTG-----D 1444
QY	2111	HYRCNLSLSDLKEIEBFHRIMKKLYVLGYVTKIKEEQHECINKPEMDFIDEVEKQKEL 2170
Db	1445	DINC--EKNDQAKESY-----LKDE-----JKKLSMYGELNKNKSY 1482
QY	2171	LIKLOHLQDCDVPSELRDOLKLNQNNMDLHIEILKDFSESEPPSIKTEFQOVLNRRKEM 2230
Db	1483	DEKYNL-----TNELKELKIR-----NKKE 1504
QY	2231	TQPLEEWLNRFDIEKLNKGIQKENDRICOVNFFNNRIIATWNEST-EFERSATISKE 2289
Db	1505	EATAE-----LNKLKN-IKEKNKSVKNDSESSNNIITKGDKTPYVSGNDKIQK 1555
QY	2290	WBODL-KSLKEKNEKLFKNYQTLKTSLASGAQVNPPTQDNKNPHVTSRATOLTTEKI--- 2345
Db	1556	WKANLVKLKKEPD-LWDNINSLE--KENFRVMSIVKENKN-----VQNDKIVGI 1602
QY	2346	-----RELENS-----LHAKESAMHKSKEIKKQKELEVTNDI-- 2379
Db	1603	YSYFKKCEKELKNDMLVICLVLDILSLFLNDNFVNLPEKIDKI--LWKQMYIPEIRI 1660
QY	2380	-----TAKLOAKVHESNKCLEKTEYIQVLQDKVALGAPKYBEETDLKMLKVLDL 2431
Db	1661	LFPIYFSFDKLRNV-----KCNVEEYVNNERYEYSWAL-FOTYLETASNLKKEIMYYVL 1715
QY	2432	ERMK-----NAKEPEKEISATKAT--VEYQKEVIRL-----LRENLRSSQAAQ-----D 2473
Db	1716	EKAEDSCENNSNFDK---PKITDILNFKDSIRLKTIAQIRKELNFEAKNINLYD 1771
QY	2474	TSVISEHTDQPSNKPLTCGGSGIVQN-TKALILK--SEHIRLEKESIKLKQONEOL- 2528
Db	1772	YQII-----LNKYHECLRKLKIVKNMARELDFNWNVSSKFSIKKELESCSDENEFK 1823
QY	2529	---IK--OKNELLSNQH-----LSNEVKTWKEK-TLKREAHQVTCENSPKSPK 2572
Db	1824	YNNIKNEESNDTIDPKPGNNLIQKIINLQRNKTEKKKNLNVNEINTWYPGDTPPG-K 1882
QY	2573	VTGTASKKKQ 2582

Db 1883 IFTTNDNSKQ 1892

RESULT 14

RESULT 14

I51617

kinesin-like protein 1 - African clawed frog

C/Species: Xenopus laevis (African clawed frog)

C/Date: 13-Sep-1996 #sequence revision 13-Sep-1996 #text_change 02-Feb-2001

C/Accession: I51617; A48835; S48837

R/Vernos, I.; Raats, J.; Hirano, T.; Heasman, J.; Karsenti, E.; Wylie, C.

Cell 81, 117-127, 1995

A/Title: Xklp1, a chromosomal Xenopus kinesin-like protein essential for spindle organization

A/Reference number: A56221; MUID:95236444; PMID:7720067

A/Accession: I51617

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 1-1226 <VER>

A/Cross-references: EMBL:X82012; NID:9562792; PIDN:CAA57539.1; PID:g562793

R/Vernos, I.; Heasman, J.; Wylie, C.

Dev. Biol. 157, 232-239, 1993

A/Title: Multiple kinesin-like transcripts in Xenopus oocytes.

A/Reference number: A48835; MUID:93246065; PMID:8482413

A/Accession: A48835

A/Status: preliminary; not compared with conceptual translation

A/Molecule type: mRNA

A/Residues: 9-162, 'L', 164-338 <VE2>

A/Residues: 9-162, 'L', 164-338 <VE2>

A/Experimental source: oocyte

A/Note: sequence extracted from NCBI backbone (NCBIP:130975)

C/Genetics:

A/Gene: klp1

C/Superfamily: unassigned kinesin-related proteins; kinesin motor domain homology

C/Keywords: ATP; nucleotide binding; P-loop

F:9-343/Domain: kinesin motor domain homology <KMOT>

F:87-94/Region: nucleotide-binding motif A (P-loop)

Query Match 7.0%; Score 935.5; DB 2; Length 1226;
Best Local Similarity 24.6%; Pred. No. 5.1e-18;
Matches 341; Conservative 261; Mismatches 478; Indels 309; Gaps 45;

Qy	3	ERG-AVAVCVVRPLNSREESIG-ETAQVYMKTDNNVIYQVDGSKSFNDRVFHGNETTK	60
Db	4	DEGIPVRVALRCRPLVPKENNEGCKMCLTFVPEQOVI--VGTESFTDYDFPSAQOE	61
Qy	61	NYVEEIAAPIIDSALQGYNGTIFAYGOTASGKTYTMGSEDH-----LGVIPRAIHI	113
Db	62	EYVNSAVAPLTKGIFKGNATVAYGQTGSGKTYSMGAYTINQENEPVGVIPRTVIAL	121
Qy	114	FQIKKFPDRFPLRVSYMEIYNETITDL-CGTQMKPLIIREVDNRNVYVADTTEEVV	172
Db	122	FREIHQRPWEFNLKVSYLEIYNBEILDLLYAARDKNTWISIREDPKGIKGLTERDV	181
Qy	173	YTESWALKWIKTGKSRHYGTEKWNQSRSGHTIFRMTLESREKEGPENCEGSKVSHLN	232
Db	182	KTALDTLSCLQGNSRRTVASTANWSSSRSHAFTTISIEQRKGDKN--SFR-SKIH	237
Qy	233	LVDLAGSERAAQTGAAGVRLKEGGINRSLFILQVIKKLSD-GOVGGFINVRDSKLTRI	291
Db	238	LVDLAGSERQKTKAEGDRLKEGISINRGLLCLGNVISALGDESKGGFVPRDSKLTRL	297
Qy	292	LONSIGNPKTRIICITIPV--SDELTALQFASTAKYMNTPVNVNEVSTDEALLKRYR	349
Db	298	LQDSIGNSHLTACVSPADSNMEETLNTLYRADRKIKNKP IVN-TDPAQAELOQLK	356
Qy	350	KEIMDLKQLEEV-----SLETRAQAMEKD-----LQAQ	378
Db	357	LOVOELQVLLQAHGGTLPVLSNMEPSENLQSLMERNKNLEKENGKLSRELGEAAVQTAQ	416
Qy	379	LLEEKDILLQVQNEKI-----ENLTRMLVTSSSLTQ-----QELKA	415
Db	417	FL-EKIIMTQQQNEKUGSKMEEIKQHAACKVNLQRLVETLEDQELKONVEIQNLQOOVIV	475
Qy	416	KRXRVRTWICLGINKMNKSNYADQFNIP-----TWNTTKTKISINLREI	461

Db 476 QLODESSGAGSIEAMDEB--AASFVPPEDESGEKSSDGFNTNHALROAQSKELI-EL 532
 QY 462 DESVCSSESVFNTLTLSEIENWPAKLLNQENIESELNADYDNLVLDYEQRLTEK 521
 Db 533 NKALVMKEALAKMAQNDQLEPIOSEYLNKHLSEVGVQKEKEBELIALHSAKKN 592
 QY 522 EEWELKLEKNDLDEFEALERTKQDQEMQLIHEISNLKLVKRE- - - - -VYNQDLE 574
 Db 593 NQAKUSERRRKLQELGEGOMTELK- - - - -KGEQSKLLKLRSTKTKVAKNQELQ 644
 QY 575 NELSSVVELLEKEDQIKLOEYIDSKLENIMKDLSSYLESTEDPKQMKQTLFDATVA 634
 Db 645 GMKQORVQLMRQMKDAEFRTWKQKTEVIQL- - - - -KEDKRQVVELL- - - - - 690
 QY 635 LDAKRESAFRSENLEKEMKELATYKQENDIQLYQSOLFEAKKMQVDLEKELQSAF 694
 Db 691 - - - - -KLERDFQKQANV-LRRKTEEAASANKELKALQKQKAMERKQDSQ- - - - -SKMEGAA 743
 QY 695 NETTKLTSIDGKVPKOLLN-LELEGKITDLQELNKEVEEALREEVILLSELKSLP 753
 Db 744 SRV- - - - -KNWLANEVEVLSTEEAQRHLNLDLEDRKILAQDI- - - - -AQLKQKT 788
 QY 754 SYVERLKEIQDKSEELHIITSEKDLFSEVVKESRVQGLLEHIGTKDDLATQSNYK 813
 Db 789 DAGERIPTKIRRT- - - - -YTVAELENLEBEASVTK- - - - - 819
 QY 814 STDQEFONFYLHMDPEQKVMVLEBNERMNOELVNLSEAKQFDDSLGALKTELSYKQ 873
 Db 820 - - - - -QIESLETMEELRSA 833
 QY 874 ELOKTEVROERLNEMLQKLENRDPSLQTVEREKTLITEKLOQTLEEVK- - - - -TLTQE 930
 Db 834 QIADLOQLKLDAGEEEMVWRWETISNIMEAKALKYLITELVSSKVAGSKLESSVKQN 893
 QY 931 KODLKQLOESLQTERDQKSDIHTVNNIDTQOLRNALSELKQHEITINTLAKISE- 989
 Db 894 RAHVADLQKNIFEERNQW- - - - -AEMETEHSQSL- - - - -MQLEQHQBKILYLLSLOQK 943
 QY 990 EVSNLMEWENTGTKEDEFOQKMW- - - - -GTDKKQDLKAKNTQTLTADVKDNEIIEQ 1041
 Db 944 QASVPVTIEELPABEITEREKQMLERLKFQDEETEKMKALCKKQQL- - - - -QENDMYKQ 999
 QY 1042 QKTFSLIQKELQOMLESVIAEKEQLKT- - - - -LKENIEMTIENQBELR 1088
 Db 1000 KLAILHVASGK- - - - -LHNILPRAEICSPPDFIPPKRGKRTNAXSAVILEDL- 1053
 QY 1089 LLGDELKQOBIVAQEKNAHAKKESGELSRDRLAEVEKL- - - - -KEKSQLOEKQOQLLV 1146
 Db 1054 LSESESEESDDKNWEPGNNSKQSKLTSKCSKARCGNKMCGCKRTKQ- - - - -NC 1104
 QY 1147 QEMSEMQKLINELENKLNKELTLEHMETELELAQKLNENYEVKS- - - - - 1197
 Db 1105 SDDCFDPSKCRNDRNHDEGHEDQSLE-SENSKID- - - - -YPDVTAGGSFTTFC 1155
 QY 1198 ITKERVKLKELQKSPETERDHLRGVIREIATGLQTKBELKIAHILKHEQETIDELRRS 1257
 Db 1156 VTPYKVLRLISD- - - - -IGVLSIKLQKPSASASVMSQEN- - - - -QTS 1199
 QY 1258 VSEKTAQII 1266
 Db 1200 ILTKKKKVL 1208

RESULT 15
 T49451
 N:Alternate names: protein B14D6.30
 C:Species: Neurospora crassa
 C>Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
 C:Accession: T49451
 R:Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,
 submitted to the Protein Sequence Database, May 2000

A:Reference number: Z25022
 A:Accession: T49451
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1742 <SCH>
 A:Cross-references: EMBL:AL356173; GSPDB:GN00116; NCSP:B14D6.30
 A:Experimental source: BAC clone B14D6; strain OR74A
 C:Genetics:
 A:Gene: NCSP:B14D6.30
 A:Map position: 6
 A:Introns: 58/1; 166/1; 267/3; 1543/3
 Query Match 7.0%; Score 929.5; DB 2; Length 1742;
 Best Local Similarity 21.5%; Pred. No. 1e-17;
 Matches 452; Conservative 351; Mismatches 683; Indels 621; Gaps 81;
 6 AVAVCVVRVP- - - - -LNSRESLSETAQVYWKTDNNVYQVDSGSFNF 49
 Db 54 AVRVAVRVPPLGPDGPDGVDLPQRFQSMVQVQGETGV- - - - -AIDSPQGRKLFVF 105
 QY 50 DRVFHGNETTKVYVEIAPIIDSAIQVNGTIFAYGQTASGKTYM- - - - -MGSE- - - - -DH 102
 Db 106 DRVF-GPEVDQEGWBYLSDCVNAFTQGYNVSLLAQGGAGKSYTGTAGPDVQEDLEA 164
 QY 103 LGVIPRAIHIDIFOKI- - - - -KFP- - - - -DREFLLRVSYM 132
 Db 165 MGVIPRAALFEKLDGSSPKSQGAASKSSQLRAPSRVAMLQPSNIDKDWKLATATYV 224
 QY 133 EYINETITDLC- - - - -GTQMKPLIIRDVNRNVVADLTVEEVYTSMAKMIT 183
 Db 225 EYINETLRLIPEHIPQHERGT- - - - -VTIRBDVKNILITGLQ- - - - -V 266
 QY 184 KGEKSRHYGTQKNOSSRSRSHITFRMLLESKEGEP- - - - -SNCEGSVKV 228
 Db 267 EGSALRQTDATANAKSSRSHAVFSLNLRVKGAKGPTAPTDRRMSMLEAMSGTEAMVT 326
 QY 229 -SHLNLVDLAGSRAAQTGAAGVRLKEGCNINRSLFILGVVVKLSQGVGGFFNYRDSK 287
 Db 327 DSMHFDVLAGESELKNTGQGERAKSGISINAGLAALGKVISQLSSRQPGAHVSYRDSK 386
 QY 288 LTRILQNSLGNPKTRIICITIPVSF- - - - -DETALQFASAKYMKNTFYNEV- - - - -STDEA 343
 Db 387 LTRLLQSLGNAITYMIACTVQAEFHLSETLNTVQARARATQKPRIQQVDEGDKQA 446
 QY 344 LLKRYKEMDKLQKLESEVSLTEAQAMEKDLAOLLEKDLQKQVNEKINENITRLMVT 403
 Db 447 IIERLKAFAFLREQIRS- - - - -SERGGGDRRNNLLAPGERSER-QNERAEALQNLQD 499
 QY 404 S--SSLTLQELKAKRRRTVTCIGKINKMKNYADQFNPTNITTKHLKLSINLLREI 461
 Db 500 ARENYTILSQ- - - - -RHAKLISEMAKARENEFAEN- - - - -OHL 532
 QY 462 DESVCSSESVFNTLDTLSEIENWPAKLLNQENIESELNADYDNLVLDYEQRLTEK 521
 Db 533 EESL-GES- - - - -ATERLNRS- - - - -NSFAQAVEQVWLEYE- - - - - 562
 QY 522 EEMELKLEKNDLDEFEALERTKQDQEMQLIHEISNLKLVKREVNQDLENELSSKV 581
 Db 563 - - - - -KT- - - - -IQSLEQSLAS- - - - - 574
 QY 582 ELLREKEPQIKKLOEYIDSKLENIMKDLSSYLESTEDPKQMKQTLFDATVALDAKRES 641
 Db 575 - - - - -TRATLANTTEATLLEKETKC 593
 QY 642 AFLRSNELKEMKELATYKQENDIQLYQSOLFEAKKMQVDLEKELQSAFNEITKLT 701
 Db 594 AYTTINTQLOARLOKLMREASTEN- - - - -YLHDLA- - - - -KLD 629
 QY 702 SLIDGKVPKOLLNLELGKITDLQELNKEVEEALREEVILLSELKSLPSEVERLRK 761
 Db 630 SHTSGEERKNAVIT- - - - -ITELRKETAR-VRENEANAEDYISTLE- - - - -ERLAE 672
 QY 762 ETQDKSEELHITTSKDKLFSVWHKE- - - - -SRVQGLLEIGTKDKODLATQSNYKSTQDE 818

Blank sheet
USPTO